DOCUMENT RESUME

ED 163 807 FL 009 924

AUTHOR	Perren, G. E., Ed.
TITLE	Foreign Language Testing. Specialised Bibliography.
INSTITUTION	Centre for Information on Language Teaching, London (England).
PUB DATE	Jul 77
NOTE	72p.
AVAILABLE FROM	Centre for Information on Language Teaching and Research, 20 Carlton House Terrace, London SW1Y5AP England
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Abstracts; Achievement Tests; *Annotated Bibliographies; Aptitude Tests; Cloze Procedure; *English (Second Language); Evaluation Methods; Language Instruction; Language Proficiency; Language Skills; *Language Tests; Modern Languages; Psycholinguistics; *Second Language Learning; *Testing

ABSTRACT

This specialized bibliography provides a widely representative selection of materials on testing foreign language skills. It is divided into two sections. The first covers articles that have appeared in periodicals over the past two years and includes abstracts of each. These have been reprinted from volumes 1 through 9 of "Language Teaching and Linguistics: Abstracts" (formerly "Language Teaching Abstracts"). The second part is a selected list of books dealing with language testing and a brief selection of published tests. A separate contents index of the abstracts is provided, and a combined index of authors and editors covering both books and articles is appended. (NCR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163807

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barbara Mewson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Published by the
Centre for Information on Language Teaching and Research
20 Carlton House Terrace
London SW1Y 5AP

© Centre for Information on Language Teaching and Research, 1977

ISBN 0 903466 16 3
Printed by Premier Copying Service Ltd. 26 Osborne Street, London E1

Specialised Bibliography

Foreign language testing

Edited by G E Perren

July 1977

Contents

		Page
1.	Introduction	5
2.	Contents index of abstracts	6
3.	Abstracts	7
4.	Books	66
5.	Index of authors and editors	70

1. Introduction

In its broadest sense testing must embrace not only standardised and objective-type tests but also examinations and all forms of systematic assessment of performance. Needless to say, this bibliography cannot claim to be completely comprehensive over such a wide field: it attempts only to provide a widely representative selection of materials on testing foreign language skills.

It has two main parts. The first covers articles which have appeared in periodicals over the past ten years and abstracts of the contents of each are provided. These have been reprinted from volumes 1 to 9 of *Language Teaching and Linguistics: Abstracts* * *(formerly Language Teaching Abstracts)*. *All* articles on language testing which have appeared since 1968 are not included, but most of those appearing in major periodicals will be found. The second part consists of a selected list of books dealing with language testing and a brief selection of published tests. A separate contents index of the abstracts is provided while a combined index of authors and editors covering both books and articles will be found at the back of this volume.

Language Teaching and Linguistics: Abstracts is a quarterly journal, compiled by the English Teaching Information Centre of the British Council and the Centre for Information on Language Teaching and Research, published by the Cambridge University Press. It provides summaries in English of selected articles appearing in about 400 journals covering all aspects of language teaching and applied linguistics.

2. Contents index of abstracts

The abstracts of articles have been arranged alphabetically under authors and then sequentially according to date of original publication.

No attempt has been made to provide a completely analytical subject index but entries are numbered and listed in a simplified contents index. Rather arbitrarily a broad categorisation of the main subjects is employed which should guide users to find what they want. A full author index is provided on page 70.

It will be noticed that individual languages do not appear in this contents index. Most articles refer to specific languages (the largest number refer to English, with French next and German third) but all are of potential interest to teachers of any foreign language, although techniques may be exemplified by reference to one only.

Aims

- Achievement testing — 12, 19, 48, 62, 102
- Aptitude testing — 18, 21, 64, 65, 69, 83, 84, 103, 111
- Diagnostic testing — 8, 27, 123
- Proficiency testing — 12, 13, 30, 37, 42, 50, 60, 71, 80, 98, 100, 108, 109

Techniques

- Cloze procedures — 1, 2, 3, 29, 57, 74, 87, 89, 101, 120
- Objective methods — 9, 10, 15, 46, 54, 59, 76
- Television — 121
- Theory and principles — 19, 20, 22, 34, 38, 58, 59, 67, 71, 73, 75, 78, 80, 82, 88, 91, 93, 94, 98, 107, 109, 112, 124, 132

Testing particular skills

- Dictation — 86
- Grammar and structure — 40, 56, 79, 99, 115, 127
- Language dominance — 131
- Reading — 1, 2, 3, 15, 17, 23, 26, 30, 44, 52, 53, 56, 60, 72, 74, 110, 114, 116, 128
- Speech
  - comprehension — 16, 25, 27, 30, 31, 36, 41, 45, 53, 56, 57, 70, 81, 97, 118, 119, 122, 126, 130
  - production — 4, 5, 6, 7, 16, 19, 31, 32, 33, 36, 45, 49, 68, 85, 96, 105, 106, 108, 110, 113, 117, 118, 119, 122, 126
- Spelling — 92
- Translation — 15, 77, 95, 132
- Vocabulary — 55, 56, 63, 66
- Writing — 15, 24, 35, 42, 47, 51, 110, 129

3. Abstracts

1. **ANDERSON, Jonathon**
A psycholinguistic approach to the measurement of reading materials comprehension *Kivung* (Boroko), **3**, 2 (1970), 83–94.

The contribution of psycholinguistics to the problem of reading comprehension measurement is studied by means of a comparison between cloze procedure, Watts' reading comprehension test and a traditional comprehension test by questions. The cloze reading test is shown to be a highly reliable measuring instrument and a valid measure of specific as well as general reading comprehension.

2. **ANDERSON, Jonathon**
A technique for measuring reading comprehension and readability *English Language Teaching* (London), **25**, 2 (1971), 178–82.

The article justifies cloze procedure as a measure of reading comprehension and readability, and describes how cloze tests are constructed for use in the classroom. The subject's task is to replace the words deleted from a passage, and the number of words he correctly replaces is an index of his comprehension of the passage. Blank-filling and sentence-completion tests are not the same as cloze procedure. Cloze procedure may be used to measure the reading difficulty of English and the comprehension abilities of non-native speakers of English. [Short bibliography.]

3. **ANDERSON, Jonathon**
The application of cloze procedure to English learned as a foreign language in Papua and New Guinea *English Language Teaching* (London), 27.1 (1972), 66–72.

The value of cloze procedure as a measure of readability and of reading comprehension in English as a foreign language was tested by means of an experiment at a primary school in Port Moresby. Several methods of scoring cloze tests were also investigated: verbatim score, synonym score, alternative response score, and grammatical class score. Two paragraphs were selected from each of three readers not in use at the school. Cultural factors meant that readability formulae were unlikely to be reliable. The passages were therefore rated in order of difficulty by teachers experienced in the teaching of English in Papua and New Guinea. [A detailed description is given of the method of differentiating between the levels of difficulty of the various reading passages and discriminating between the abilities of the subjects of the test. Results are tabulated, showing: (1) the mean percentage passage-score standard deviations, judges' ranking and cloze ranking of the reading difficulty of the passages, (2) the relative difficulties of the passages under the five scoring procedures, and (3) the correlation matrix for verbatim, synonym, alternative response, and grammatical class scores.]

The cloze technique ranked the passages in the same order of difficulty as did the teachers, and discriminated between the abilities of subjects. The various scoring procedures differed little, the simple verbatim method being the most objective and least tedious. [Short bibliography.]

4. **BACHMAN, Lyle F**
Testing oral production *Bulletin of the English Language Center* (Bangkok), **3**, 1 (1973), 41–58.

An individual's proficiency in oral production is affected by his linguistic competence, by psychological and environmental factors, and by the social and semantic construct of the situation. Since we cannot directly test linguistic competence, any measure of the individual's oral production is a measure not only of his competence but of the non-linguistic factors affecting his performance. It is necessary to re-define the criteria for measuring oral production in functional or situational terms.

The 'discrete point' approach to the testing of oral production assumes that a number of discrete elements can be identified. The assumption that the sum of the discrete elements equals the individual's overall proficiency is questionable. The best type of oral production test is one in which proficiency levels are defined in functional terms, and which attempts to test complete oral proficiency and overall oral communication. The semi-structured oral interview is still the most valid test in ideal circumstances, but interviewers have to be trained and uniform grading standards are hard to maintain. Another approach involves the elicitation of highly structured speech samples [various objections to this technique are raised]. The use of structured dialogues to which the student must respond in specific ways is a comparatively reliable method. Techniques involving the use of pencil and paper are marginally valid.

Recent research has focussed on global techniques [some recent experiments are described]. Discrete-point tests help with diagnosis of specific learning problems and are more reliable than global methods, which are more in keeping with research in linguistics and psychology, and capable of testing factors not measured by discrete-point methods. The ideal test will probably contain both these techniques.

5. **BEARDSMORE, H Baetens**
Testing oral fluency *IRAL* (Heidelberg), **12**, 4 (1974), 317–26.

Discrete-point tests do not measure performance in real situations, and the scoring of interviews is subjective and variable. It should be possible to combine the two types of test: both linguistic and communicative ability should be tested. A test of oral fluency adopted after various experiments is described. Students record briefly on tape, in the language laboratory, what they have to say about a course-orientated topic, chosen from three [examples]. Throughout the year they are given trial tests of this kind. Fluency, accuracy, relevance, intelligibility, pronunciation, variety of

structure and of lexis are the scoring criteria, and there is remarkable consistency among the various markers. Although there are various unsolved problems and further research needs to be undertaken, the experiment has called attention to some of the elements involved in interview testing and has increased the element of objectivity in such tests. [Bibliography].

6. **BEARDSMORE, H Baetens, and A Renkin**
A test of spoken English *IRAL* (Heidelberg), **19**, 1 (1971), 1–11.

One of the most difficult fields in language testing has proved to be that of productive spoken language. This has been one of the foremost considerations at the Institute of Phonetics of the University of Brussels where an objective test of spoken English has been designed. Testing methods are reviewed: the 'strip-cartoon' technique, the 'interview' technique, and the 'dialogue' technique. The latter was chosen for this particular test. The students were led to participate in a dialogue with the voice presented on the tape-recorder. Items were carefully selected to leave very little scope for variations in the answers, and validity was tested on a wide variety of native speakers including very young children. Instructions were given in the native tongue before each section. In correcting, a strict principle of 'right' or 'wrong' was adhered to. A printed correction sheet was used giving the most likely answers (including alternatives) and leaving room for variants to be written in. Incorrect variants provided useful information for error analysis. A supplementary test was added to give students the opportunity to speak more freely. The general conclusion is that at a given level of language ability the information provided was sufficient to replace the interview technique.

7. **BEAUDOT, Alain**
Le 'SAID system' *Langues Modernes* (Paris), **61**, 4 (1967), 395–8.

The 'Speech Auto-Instructional Device', developed at the University of Michigan by Harlan Lane and his colleagues, is used for analysing and teaching the prosodic features of a speech pattern. Because of its high cost the machine is more useful at the moment for providing psychological data, as described by the teleprint records of the student's learning behaviour, than for teaching.

The machine examines automatically and separately the efforts of the student to attain an acceptable standard in the pitch, amplitude, and tempo of the patterns being learned. The student is allowed to pass to a second feature only when he has mastered the first, which may require various repetitions. Each time the student imitates the model sentence his efforts are recorded by the teleprinter, which provides a picture of his progress. Different standards of acceptability by the machine can be pre-set, according to the degree of perfection sought. In spite of one or two points of regression in the student's performance it is found that he

always reaches his goal. It has been found that the machine is much more accurate than skilled linguists in the assessment of performance in the three aspects mentioned.

8. **BLATCHFORD, C H**
A theoretical contribution to ESL diagnostic test construction *TESOL Quarterly* (Washington, DC), **5**, 3 (1971), 209–15.

A proficiency test is norm-referenced, involving comparison between one student's performance and that of others: a diagnostic test should be criterion-referenced, aimed at indicating whether content has been mastered. The degree to which a test is diagnostic depends less upon the purpose of the test than upon the way in which the scores are analysed. A total score may be interpreted as a measure of proficiency, but part scores serve a diagnostic purpose by indicating particular strengths/weaknesses. Indeed, a diagnostic test should have no 'total' score: it should consist of a series of miniature tests on specific problems.

It is believed that reliability depends upon the length of a test. Often, however, practical difficulties militate against the construction of a long test. To examine the reliability of miniature diagnostic tests an experiment was conducted based on a forty-item test of ten grammatical points, each point tested by four multiple-choice items [examples given]. Reliability figures were calculated on the basis of responses to all four items, to three items and to two items. The conclusion is that for diagnostic purposes a two-item test of a grammatical point is as reliable as a four-item test, provided that test makers follow the usual canons of carefully constructing and pre-testing items. Time may thus be saved in constructing tests, but more time should be devoted to analysing test results to establish the range of remedial instruction required.

9. **BONHEIM, Helmut**
Objective testing in English studies *Neuere Sprachen* (Frankfurt am Main), **19**, 4 (1970), 191–200.

No single test can cover all the different capacities needed for mastery of a foreign language. In a conventional examination it is very hard to test the pupil's grasp of small units of knowledge. An objective test can include many more questions and will give a more complete picture of what the pupil knows. [Example of tests covering literary history, language skills-synonyms, critical terminology, figures of speech, multiple choice comprehension questions on a text]. Marking standards vary considerably with the conventional, translation type of foreign-language examination. Objective tests make possible rapid and uniform marking and more time is left for marking essay papers. Wherever a small sample can fairly be expected to mirror accurately a large area of knowledge, objective testing can be fairly used.

10. **BREITUNG, H A**, and others
Zum Problem des Tests, insbesondere des Einstufungstests, im Deutschunterricht für Ausländer [The problem of testing, especially for grading students into courses of teaching German to foreigners] *Deutsch als Fremdsprache* (Leipzig), **11**, 6 (1974), 342–53.

Hitherto most so-called objective tests have contained some subjective factors; objectivity is harder to secure when, as with language acquisition, the matter to be tested is complex. Until the grading tests were devised students had been assessed solely on an oral interview; this practice had been unreliable. Now the oral interview is followed by a test appropriate to the four stages: beginners, basic, upper, advanced.

The aim is to grade more objectively and accurately and the tests have proved reliable in assessing knowledge and receptive skills; their least exact part is the testing of oral skill. There are three written tests, each including 15–20 items of grammar, vocabulary and word formation, five items of comprehension through hearing, together with some spelling and punctuation. The pass mark is 35 out of 50. The 1971/2 tests showed a pass rate of 51 per cent [analyses of performances in various items of the tests]. [References.]

11. **BRIÈRE, Eugène J**
Current trends in second language testing *TESOL Quarterly* (Washington, DC), **3**, 4 (1969), 333–40.

[This paper, given at a conference on 'Problems in Foreign Language Testing' in 1968 attempts to summarise the content of some of the papers presented at that conference.] Papers presented by Spolsky, Upshur, and Jakobovits were mainly concerned with identifying the variables which define the competence of a speaker or listener. Plaister discussed 'Measuring control of English structure of non-native speakers of English by means of responses to a scale of grammaticalness'; Palmer's paper was on a 'Proposed analysis of some factors affecting the rating of English spoken by foreign speakers of the language'; Wilson spoke on 'ESL materials for the Navajo'; Metfessel on 'Assessment of cognitive abilities of children and youth from the culture of poverty'; Brière on 'ESL placement tests for the Bureau of Indian Affairs'; Youseff on 'Facilitation effects of multiple choice items in grammar testing'; Harris on 'Testing English in difficult circumstances (Greece)'; and Jones and Kaplan on 'A cloze procedure test of listening for university level students of English as a second language'.

12. **BRIÈRE, Eugène J**
Are we really measuring proficiency with our foreign language tests? *Foreign Language Annals* (New York), **4**, 4 (1971), 385–

Definitions of 'achievement' and 'proficiency' in foreign language performance are given. The development of different methods of teaching foreign languages and the tests related to each method are reviewed and

discussed. Current attempts to measure 'proficiency' in foreign languages are considered inadequate. A plea is made for increased interdisciplinary experimentation in all areas of language testing.

13. BRINGS, Friedhelm
Zur Konstruktion eines informellen Sprachleistungstests [The devising of an informal language proficiency test] *Praxis des neusprachlichen Unterrichts* (Dortmund), **17**, 4 (1970).

It is well within the scope of any individual teacher to set up an informal language proficiency test for the pupils he teaches. He must decide carefully which skills he wants to test, whether there should be a time limit for any individual test items and what types of question he should use. [Examples are given of a number of types of question, and each type is discussed in some detail.]

Based on his own experience with short-answer and multiple-choice items in tests that he set up for a beginners' class of German pupils learning English, the author discusses the special difficulties involved in deciding on the questions. The instructions given to the pupil should contain nothing that is irrelevant or ambiguous to confuse him. The questions themselves should, above all, be unambiguous and deal with one language problem at a time. In multiple-choice questions the incorrect choices must be definitely wrong, yet appear plausible to the pupil. However, deliberately setting traps should be avoided.

Although devising an informal language proficiency test may be time-consuming it makes objective evaluation of a pupil's work much easier, and the test can always be re-used.

14. BROSCH, Gerd
Das Zertifikat 'Deutsch als Fremdsprache' [The certificate 'German as a foreign language'] *Zielsprache Deutsch* (Munich), **1**, (1972), 1–5.

The examination and certificate date from 1971. They were designed to narrow the understandable gap between the aims of reform movements in modern-language teaching and the achievements of language classes in the educational system as a whole. The language proficiency of school and college leavers was not up to the requirements of the period. The *Volkshochschule* was alone able to provide facilities for further education of the appropriate kind, but needed to define the goals of its instruction. A committee established a minimum vocabulary and structure list for English, the first language tackled. Since 1968 certificates have been offered for English, French, Spanish, Russian and German as a foreign language. The introduction of German as a foreign language in German colleges led to the setting up of a minimal word list and a list of high-frequency grammatical structures. These, together with a definition of goals and a description of the methods of examination, was published in a brochure in the middle of 1972. [Address given.]

The examination tests knowledge sufficient for coping with everyday situations. The emphasis is on spoken language and on the active use of language. Objective assessment is used and the examination is well defined. The certificate is intended to denote a basic knowledge, although an extension along lines already taken by the English certificate is not out of the question. It can also offer a standard of proficiency in written and spoken everyday German, and a basis for different courses in German.

15. **BUBENÍKOVÁ, Libuše**
Formy jazykových testů [Language-testing techniques] *Cizí jazyky ve škole* (Prague), **15**, 5 (1971/2), 198–207; **15**, 6 (1971/2), 253–64; **15**, 7 (1971/2), 298–307.

Authors of foreign-language achievement tests must consider three basic questions: which aspects of the language are to be tested, which techniques or forms should be employed (recognition or production testing), and which material is to be built into the test.

Response-item techniques (recognition testing) include the choice between alternatives, the correct ordering of items according to meaning, reading aloud (not allowing for group testing), dictation (partial dictation, ie filling the blanks in a text is considered useful for the regular assessment of progress), dictation and its translation (only recommended for testing interpreters), texts with inserted words (valuable to test reading comprehension speed in the foreign language at an advanced level only).

Multiple-choice techniques deserve special attention in foreign-language recognition testing. In spite of certain drawbacks this form has proved useful in testing comprehension through reading, comprehension through listening (preferably combined with visual pointers), often including sound discrimination, vocabulary resources (by means of synonymy, antonymy, definition, paraphrase and substitution tests) and grammar (proposed answers should not include incorrect forms). Authors of multiple-choice proficiency tests should have a clear idea of what they want to test. Separate language aspects should not be combined in one test (eg vocabulary tests should not include features of spelling or phonemics), nonsense forms should be avoided (eg *writed*), if possible all tested problems should be referred to in the context of a sentence (or pictorial situation), all choices in each item should be presented in a uniform manner, no visual indication of the correct choice should be implied at first sight, and in a test battery multiple-choice items should alternate with other techniques.

Production testing (free response items) can be divided into objective free response items (only one independently encoded answer is possible), subjective composition (the essay type), tests of written and oral expression and translation tests. Objective free response tests may include substitution tests (eg stimulated by one word), completing the end of sentences and forming enlarged sentences on the basis of a nucleus of words. In tests of oral and written expression students should be set a well-defined task based on a pictorial framework (limited in content) on

a dialogue, and on the reproduction of a spoken or written text. In translation tests partial translation is favoured, ie the student is asked to translate only parts within a given framework of a foreign-language text.

16. **BUTZKAMM, Wolfgang**
Mündliche Tests im elementaren Fremdsprachenunterricht [Oral tests in foreign-language teaching] *Englisch* (Berlin), **6**, 1 (1971), 5–9.

Pupils are unaccustomed to oral work having much effect on their end-of-term or examination marks. This has its effect on the teacher as much as on the pupils. Unless more emphasis is laid on the oral tests at an advanced stage, speech will be encouraged only in the elementary classes. [Suggestions are made for types of oral test suitable for senior pupils.] The administration and scoring of language laboratory tests is, however, known to be time-consuming for the teacher and such tests could probably only be given twice in a school year, but if they are given at regular intervals they will acquire greater value in the eyes of pupils and parents. Objectivity is also difficult to achieve, but it can be done by means of tests of recognition, and correct comparison of sounds. The ideal is to test only one thing at a time. It is helpful if the pupils are given a sample list first and the correction procedure is explained to them. Question-and-answer conversations on a particular theme can be tested for (1) sense, (2) choice of words and structures, (3) amount of information volunteered in the reply. In this case pronunciation would not count so long as the reply was understandable. Oral teaching must never be allowed to degenerate into teaching for tests. Efficient teaching with slightly unsatisfactory testing is infinitely preferable to the reverse.

17. **BUTZKAMM, Wolfgang**
Über Lernzielbestimmung und Leistungsbeurteilung im Englischunterricht an der Gesamtschule [Determining goals of learning and judging achievement in English lessons in the comprehensive school] *Englisch* (Berlin), **7**, 1 (1972), 8–15.

Comprehensive schools are still feeling their way, and the movement for change in long-established styles of education is affecting all aspects of work including the teaching of English. The general goals of learning are as yet ill defined, but communication and social behaviour are of prime importance. The treatment of culture, social questions and social relationships in textbooks for foreign-language teaching needs re-examining. Social interaction plays a vital part in the learning process, reinforcing aural-oral skills. The next step is to establish a hierarchy of learning goals which are objective and attainable. Special attainment tests are necessary, and this presupposes strict planning and control of the teaching material. [An example of one school's attainment report is considered in detail.]

The concept of achievement is much discussed in the comprehensive school. Competitiveness is rejected as antisocial, and the idea of co-

operative achievement is considered but this requires an 'absolute' standard of measurement. It has been tried in one area in one school. In silent reading, for example, levels of ability are laid down and progressive learning developed. In the first phase it is closely bound to oral work. Then additional passive vocabulary is introduced in self-explanatory contexts. Later, literary structures are added. There is constant self-testing, so students can work independently. [Detailed sample given.]

18. **CARROLL, John B**
Implications of aptitude test research and psycholinguistic theory for foreign-language teaching *Linguistics* (The Hague), **12**, (1973), 5–15.

Research in foreign-language aptitude testing has disclosed at least three more or less independent types of abilities that play a role in controlling rate of learning. Phonetic coding ability is the ability to identify, and store in long-term memory, new language sounds. Grammatical sensitivity is the individual's awareness of grammatical functions and syntactic patterning. Inductive ability is the ability to notice and identify patterns of correspondence involving syntax or semantics. While it appears that these abilities are not easily modifiable, foreign-language teaching may require specific techniques in the diagnosis and remedying of learning problems. Application of these findings to two contrasting theories of second-language acquisition is discussed; both deductive and inductive teaching methods are found to be important. [References.]

19. **CASTIGLIANO, Luigi**
Il controllo della preparazione [Assessment of preparatory work] *Scuola e lingue moderne* (Modena), **7**, 9 (1971), 212–22.

The teacher's aim must be to teach; pupils should learn, not simply receive marks. Language teaching at the lower secondary level should: (1) help towards the child's general education; (2) provide him with the basic skills of comprehension, speaking, reading, writing. Testing provides the teacher with feedback on a class's progress, and therefore on the efficacy of his own teaching – the real point of testing. Work must be repeated when results show that the class, or part of it, has not absorbed a particular structure. An individual's failure to make progress cannot be ignored but must be analysed and remedied, just as under compulsory school requirements each child must be promoted in due course. Any failure indicates that more efficient schooling is needed. Intensive drilling is necessary and should cover a given amount of ground at a steady pace.

Outdated teaching methods (which should be discarded) allow only for traditional testing methods, eg grammar-translation; passive learning gives scope only for comprehension tests. Modern active oral techniques are adaptable in themselves to testing, and can be graded for difficulty and variety and used for linguistic or content value. Marks need not be awarded in the early – or more difficult – stages in order not to

intimidate the pupils; encouragement and reinforcement of positive motivation should be the guiding principle. It is often sound practice to award a mark for group work. Various activities can be fitted into a one-hour lesson, cooperative or individual, and tests done on each. This increases motivation and satisfies pupils and their families. Of testing methods, traditional ones should be avoided; more 'scientific' and objective methods are preferable, including test batteries, with simple or multiple-choice options. These, however, can be difficult or lengthy to prepare. Aptitude tests are useful at the beginning of a year to gauge the homogeneity of a class. The results of end-of-year examinations, accompanied by aptitude or IQ tests if required, can be worked out on a graph mark-scheme [example shown] and the mark-range adjusted for particular purposes.

For the *licenza media* examination, written papers are inconclusive because of the chance element, and interviews are not always practicable. Objective tests are to be preferred. Since, at the end of their brief secondary school course, pupils are not in a position to have a command of the target language, it might be better not to have a public examination at that stage. While reform is pending, the teacher is best advised to relate his tests closely to his class work.

20. CATFORD, J C
Some aspects of linguistics in language testing *English Teaching Forum* (Washington, DC), **6**, 5 (1968), 16–19.

Command of all levels and varieties of language can be tested. Among the functions of tests in English are: (1) assessment of attainment for award of a qualification or for selection and placement, (2) prediction of future progress, (3) measurement of the value of teaching methods and procedures – essential in designing experiments in teaching, (4) diagnosis of individual or group difficulties, (5) measurement of aptitude. Testing usually measures performance rather than competence, but this can also be measured. One can test production or reception of language, and one can test phonology, graphology, grammar and lexis, the relations between the latter two, and features of situation. In addition the student's ability to identify dialects and registers can be tested. [Details of the kinds of test are given with illustrations.]

21. CHASTAIN, Kenneth
Prediction of success in audio-lingual and cognitive classes *Language Learning* (Michigan), **19**, 1/2 (1969), 27–39.

An experiment conducted with six classes of students, three taught by audio-lingual methods and three by cognitive methods, showed a high correlation between achievement and past academic record. For the cognitive group, the significant predictors of achievement were aptitude and school rank: the same data as would be used for other subjects on the curriculum. Prediction was less sure in the audio-lingual classes. Results

also showed that it would be possible, taking account of a student's capabilities, to place him more suitably in the one type of class than in the other.

22. **CLARK, John L D**
Foreign language testing in the United States *Audio-Visual Language Journal* (London), **9**, 1 (1971), 39–45.

The basic categories of testing are: aptitude, classroom achievement, school entrance or placement, teacher certification, special purpose, and miscellaneous.

[The author gives examples of what is available in the United States in each of these categories.] Multiple-choice techniques are widely used. Active production tests of reading and writing have also been developed. Electronic aids are widely used. Language competence has been divided into discrete areas: listening, speaking, reading, and writing.

23. **CLARKE, Mark A**
Arabic distractors for English vocabulary tests *English Language Teaching* (London), **27** (1972), 77–9.

The validity of objective vocabulary tests depends upon the plausibility of the distractors. [The author discusses tests which use both English distractors and words which resemble Arabic words.] The most practical type of Arabic/English distractor consists of an Arabic translation of the correct answer, expressed by an English word similar to the translation. A test of this kind was given to Arab students at the English Language Institute of the American University of Cairo, and the more advanced students chose the Arabic/English distractors more often than did the less advanced.

The teacher should bear in mind student confusion between native and target language when compiling objective vocabulary tests, and where English is the students' third language, items from another language can be used. This approach offers data on the students' thought processes.

24. **COHEN, Arthur M**
Assessing college students' ability to write compositions *Research in the Teaching of English* (Urbana, Ill), 7, 3 (1973), 356–71.

A research project on the assessment of English compositions in junior colleges recommends the involvement of teachers in the design and conduct of the project. At an initial workshop a score sheet was devised as a key to marking, listing 19 categories to which yes/no answers could be given. Multiple blind scoring was used to eliminate bias on the part of teachers, who carried out the experiment over two semesters; each student wrote two compositions, one before, and one after instruction [details].

Analysis of results showed that the design could be used to assess change in students' ability to write compositions.

25. **CONNELL, Peter A**, and **James G Snider**
Aural English examinations in West Africa *Educational Research* (London), **12**, 3 (1970), 235–9.

In 1965 the West African Examinations Council began investigating the possibility of introducing an Ordinary Level compulsory oral English test. Elements were isolated for an objective, standardised test which, it was hoped, would influence teaching and raise the standard of spoken English in West Africa. This test would be compulsory at Ordinary Level examinations in the early 1970s. [Details provided of the second trial testing of the new aural English examination administered to some 300 children in Lagos.] Two validity studies were done during the second trial with children from the American International School in Lagos. If the test proves successful it will need to be supplemented by a textbook for students and a teacher's book to suggest ways of inculcating good aural and oral English habits in the classroom. [Bibliography.]

26. **COOPER, Malcolm D**
Measuring language problems and attainment *English Language Teaching* (London), **25**, 1 (1970), 72–6.

The full version of a test administered at the beginning and end of a Tanzanian secondary-school course in grammar and composition is given together with some scores. Part 1 tests the pupils' ability to understand a story; part 2 their ability to produce structures in re-telling the story.

27. **CORY, Mark E**
A neglected aspect of foreign language testing: structuring the review session *Foreign Language Annals* (New York), **7**, 4 (1974), 417–20.

Considerable attention has been paid to the design, administration and interpretation of foreign-language tests, but practically none to the process of reviewing (revising) for these tests in group-paced instruction. The implementation of a fully articulated practice examination modelled on the format of the exam itself can structure a review session in such a way as to exploit more fully the teaching potential of an exam situation. Incorporating many of the self-evaluative benefits of individualised instruction, such a review sheet is helpful in co-ordinating a multi-sectioned course with common examinations, and has the additional virtue of being highly popular with students. [Sample review sheet and practice exam for German in the third semester, and discussion of their application in class.]

28. **COSGRAVE, Desmond P**
Aural comprehension tests and how to prepare them *Modern English Journal* (Ashiya-shi), **1**, 1 (1970), 5–16.

Various factors must be taken into account when preparing aural comprehension tests on grammatical structure. Meaning will play an important role in the tests, though material to be tested must be described in terms of grammatical patterns already encountered in class. The type of test envisaged is recorded for use in a language laboratory. The test is best presented as a multiple choice, showing, by a completing word or correct response, whether the structure has been understood. Choice of a correct sentence may show whether a passage presenting a situation orally has been understood. [Illustrations.] Modifications can be employed which introduce visual aids. If students who have made some progress in writing are asked to write answers to problems given orally this can also be useful, but students' chosen responses should be short. [Guidance is given on preparing the test script, recording the test, and evaluating it.]

The skills of understanding and speaking are closely related but so far no adequate test of speech capability has been devised. Nevertheless an aural comprehension test together with the teacher's assessment of his pupil's speech should give good evidence of the student's abilities.

29. **DARNELL, Donald K**
Clozentropy: a procedure for testing English language proficiency of foreign students *Speech Monographs* (New York), **37**, 1 (1970), 36–46.

The new test employs a variation of 'cloze' procedure and an 'entropy' measure derived from information theory. (The cloze procedure was used previously to measure readability, comprehension and language aptitude; the entropy measure indexes the compatibility of an individual's responses with those of a selected criterion group.) Clozentropy uses a simple method of constructing items for objective testing and a complex, mathematically precise, scoring system which avoids the right–wrong judgement on an item-by-item basis. [Detailed description given.] Administration of the clozentropy battery takes about two hours, half the time required for the TOEFL test. Computer assessment is essential for scoring. [The article reports a study conducted to test the reliability, validity and practicality of the test.]

30. **DAVIES, Alan**
The English proficiency of overseas students *British Journal of Educational Psychology* (London), **37**, 2 (1967), 165–74.

The article reports research on the language difficulties met by overseas students in Britain and on the construction of prognostic tests of English for assessing their ability to follow courses in British colleges and universities. A battery of seven tests of receptive skills, covering phonemic discrimination, interpretation of intonation, listening comprehension,

reading comprehension, reading speed and grammar, was constructed and tried out on samples of 496 overseas students in Britain, 238 overseas students in their own countries and (as a control group) 267 native English-speaking students. As criteria of the validity of the battery, students' grades or examination results and teachers' estimates of their students' proficiency were used. Except in tests of listening comprehension, reliability coefficients were all above 0.85. Factor analysis suggested principal components of segmental listening, textual listening and a rather more diffuse one of reading comprehension. Comparative figures of performance on the tests showed that of ten home language subgroups, Semitic, Mediterranean, West African and East/Central African students ranked lowest. Observation suggests that one proficiency standard in English may be accepted as adequate for success at varying academic levels of study in Britain, although this requires further experimentation.

31. **DENHAM, Patricia A**
Tests of aural-oral control of language *Papua and New Guinea Journal of Education* (Sydney), **6**, 3 (1969), 11–16.

Diagnostic tests for students learning English are useful when the difficulties faced by a group are not known. Remedial courses can be drawn up according to the errors made in the test. Tests to measure the extent of learning are often limited to what has been taught and practised over a short period. The article is chiefly concerned with tests of this second type to measure the control of aural/oral aspects of language. Aural aspects are covered by tests of perception, recognition and comprehension while tests of production correspond with oral work. Competence in one aspect does not automatically lead to competence in the other, though the two skills have sometimes in the past been considered as identical, particularly with regard to phonology. Aural tests are the best developed [details and examples of tests of aural control of vowel and consonant phonemes, of stress and intonation, and of comprehension of a passage] . The most objective tests of language production are at the phonological level and objectivity is achieved by testing only specific aspects. Tests of sentence structure and syntax are not objective and so far no technique has been devised to overcome the limitations of the traditional approach to these aspects.

32. **DEYES, A E**
How well can our students speak? (language laboratory testing) *English Language Teaching* (London), **26**, 1 (1971), 31–7.

The author describes and criticises the construction, administration and making of a series of experimental tests of students' oral production in the language laboratory

33. **DI CRISTO, Albert**
Présentation d'un test de niveau destiné à évaluer la prononciation des anglophones *Revue de Phonétique Appliquée* (Mons, Belgium), **33/4**, (1975), 9–35.

A proficiency test for an objective evaluation of the pronunciation of advanced English-speaking students is presented. [Brief survey of different techniques.] The student is given a list of items to read without a model of any kind. The list is based on an inventory of the mistakes made by the student when speaking French. Each item aims at testing a single difficulty presented in different phonetic and prosodic contexts. This test includes three different steps: phonemics, phonetics (vowels, consonants, co-articulation problems) and prosody (stress, rhythm, intonation). In order to reduce subjective evaluation, precise instructions are given to test each category. By a simplified summation of the different scores, the teacher can draw an accurate outline of the pronunciation of each student. The aim of the test is to determine objective levels for forming homogeneous groups.

34. **DIMITRIJEVIC, Naum R**, and **Dušan Djordjević**
The reliability of the subjective assessment of the pupils' pronunciation of English as a foreign language *IRAL* (Heidelberg), **9**, 3 (1971), 245–65.

Methods for an objective assessment of a learner's achievement are constantly being improved, but in spite of a substantial body of knowledge about language testing there are still areas in which the teacher has to use traditional and subjective ways of assessing the learner's achievement. These are: speech, written composition and pronunciation. A small-scale experiment was conducted to compare subjective marks given by different categories of examiners and to examine the range of disagreement in assessing the same material. There were four categories of examiners: secondary-school teachers, phoneticians, native speakers of English, and senior university students of English. Taped material was assessed and the resulting marks were analysed. The native speakers were the most consistent in their subjective evaluation of speech. The teachers proved to be the most lenient in judging a continuous text. [Tables of test results appended.]

35. **DOBLE, Gordon**
How free should the composition be? *Modern Languages* (London), **55**, 1 (1974), 6–11.

Poor performances in the traditional discussion-type essay in 'A' level French papers suggest that the aims and limitations of this activity should be established more precisely, and that alternative ways of testing pupils' ability to write more freely in a foreign language should be explored. The problems of assessment are discussed, particularly the value placed on linguistic as against communicative competence (40 per cent of marks should be given for linguistic accuracy and 60 per cent for content and

other linguistic features).

It is suggested that pupils might be less inhibited by a more controlled exercise such as the 'structured essay', in which stimulus is provided in visual, textual or spoken form to enable pupils to write cogently on a defined topic. [Discussion of teaching techniques from the early stages of greater control to eventual free-writing, and comparison with the linked technique of guided conversation.]

36. **DONNERSTAG, Jürgen**
Differenzierung im Englischunterricht auf der Grundlage lernziel-orientierter Tests [Differentiation in the teaching of English on the basis of objective-oriented tests] *Der fremdsprachliche Unterricht* (Stuttgart), **8**, 3 (1974), 19–29.

Increased emphasis on oral communication ability requires greater concentration on testing speaking ability in the teaching of English. A qualitative test oriented towards a learning target or criterion is favoured over the norm-referenced or standardised test. The basic condition for such a test is the operationalised description of learning targets. Such a description must be based on an exact determination of the behaviour by which it can be proved that a pupil has reached the objective. Learning objectives formulated in an abstract way have to be translated into concrete behaviour, because vaguely formulated aims are unsuitable for testing purposes. Difficulties are encountered when trying to predict the pupils' speaking behaviour, therefore the discrete-point approach is recommended, dividing 'speaking' into components, for each of which exact objectives can be set. Opinions vary as to whether speaking skills can be satisfactorily tested at all. It seems that complete operationalisation cannot be achieved; instead, an expressive performance objective is put forward, together with a scale of assessment factors [details].

A classification of learning objectives with five classes by Valette and Disick is described and compared with other systems. [Examples; lists illustrate the determination and classification of learnings objectives.] Taking into account the qualitative differences in objectives, graded courses can be devised for learning groups. With the help of taxonomy, differentiation in the teaching of English can be systematic, comprehensive and long-term.

37. **DUŠKOVÁ, Libuše**, and **Eduard Beneš**
K otázce objectivního zjišťování úrovně znalosti cizího jazyka potřebné pro čtení odborných textů [Finding the level of foreign language knowledge necessary for reading professional texts] *Bulletin of the University of 17 November* (Prague), **2**, (1967), 17–33.

To find the level of language knowledge in various classes and to determine the efficiency of teaching methods, it is necessary to work out an objective method of evaluating the students' knowledge.

The article reports an experiment to find the level of foreign-language knowledge necessary for reading professional texts. A test sample was constructed from groups of sentences selected from various professional texts. The groups were classified according to their difficulties into five grades. Each grade contained a test with four sentences of the same lexical and grammatical difficulties. [The translations of the five-grade test into English and German are given at the end of the article.] A complete five-grade test is presented at the beginning of the year. In further teaching stages the first grade is left out.

The test vocabulary was chosen from minimum general-science word lists compiled by the Department of Foreign Languages of the Czechoslovak Academy of Sciences.

Experience has proved the validity of the tests. They will show a student's ability in reading and understanding foreign-language texts, and satisfy the inquiries of examiners in various departments. Similarly constructed tests can also be used in secondary schools.

38. **DÜWELL, Henning**
Die Integration von Tests in Fremdsprachenlehrwerken [The integration of tests in foreign-language textbooks] *Praxis* (Dortmund), **20**, 2 (1973),, 121–7.

The question of how far tests can be integrated in a foreign-language textbook depends on the aims and structure of the book. The aims must be detailed and capable of achievement, so that end behaviour can be measured. One test may fulfil two purposes but the results must be differentiated. General aptitude tests given before starting a foreign language act not as eliminating but as prognostic tests. Aptitude tests for individual languages would be valuable. Achievement tests can be suitably distributed through the book and should complement the teacher's judgement. Ideally tests should increase motivation. The choice of exercise or stimulus will depend on the skills to be tested, but questions of relevance, degree of difficulty and effects on the learning process must be borne in mind. The advantages of integrated tests are a unity of method in teaching and testing, regular checks on progress and early diagnosis of difficulties.

39. **EWEN, Elizabeth** and **Caroline Gipps**
Tests of English for immigrant children *Multiracial School* (Bath), **2**, 2 (1973), 22–4.

The National Foundation for Educational Research constructed diagnostic and proficiency tests of English, for use with immigrant children in primary schools. It was felt that content-referenced tests were most suitable (measuring what the pupil can do in terms of the skills being tested, rather than in comparison with his peers). The tests are divided into the four skills – listening, speaking, reading and writing – and are arranged in three levels of complexity. [Description of the type of material used at

each level.] It is hoped that teachers will be able to use the tests to diagnose a child's proficiency in various skills and on separate grammatical or lexical points. [Receptive skills (listening and reading) are scored objectively while a global approach to scoring productive skills (speaking and writing) is felt to be more appropriate.] The tests are being used as the basis of a nation-wide survey sponsored by the Department of Education and Science.

40. **FAMIGLIETTI, Michele**
Sperimentazione sulle prove grammaticati di costruzione guidata [Experimentation with grammar tests following lessions which use guided construction] *Rassegna Italiana di Linguistica Applicata* (Rome), **3**, 1 (1971), 29–53.

The tests were carried out in the scholastic year 1969–70 on a group of twenty-five boys, aged between eleven and twelve. The students were from a southern Italian working-class background, whose families were attempting to establish middle-class urban roots and habits. Their normal speech was a neapolitan dialect; the national language was heard through the mass communication media and was used only in official situations; the majority were unable to express themselves fully in it. Intelligence tests indicated that they were a normally able group, and corresponded in cultural and intellectual ability to their national and American counterparts.

'Guided construction' is a type of grammatical exercise based on the following precept: the student is asked to compose a linguistic sequence according to the specific demands of a grammatical order, using the basic expressive elements given him. For example: Maria, to approach, friend, street, becomes 'Maria approached her friend in the street'. A varied scale of examples of guided construction can be developed from the initial module.

As it is possible to demand from the student variations of construction in grammatical order, likewise the provision of expressive elements offered to the student can be varied: these can range from simple expressive words to those related to an abstract concept or a situation, or even proceed to other fields of perception, eliminating words and substituting photographic images.

The plan of the tests was to start from syntax and move towards an analysis of morphology, phonology and other linguistic elements as functional parts of the sentence. The forms of the tests reflected, where possible, structures with which the students were already familiar, and were carried out during the year simultaneously with normal work in Italian. The objectives were to measure the didactic value of guided construction; to delineate a typology for guided construction; to determine its functional aspects; to determine the efficacity of each of these types and their values in relation to other types of grammatical exercise; and to develop linguistic and expressive powers.

Results indicated that students performed better on tests of simple construction, in both written and oral free composition. Tests based on photographic rather than verbal stimulus gave more rewarding results than those based on word/situation or word/concept, which in turn showed better results than those based on lexical features. This implies that the freer the student is in his mode of expression, the greater become his powers of reflection and expression. The author suggests that guided construction could usefully be used for tests and courses on syntax.

41. **FARRINGTON, Brian**
A computer-aided test of aural comprehension *Audio-Visual Language Journal* (London), **9**, 2 (1971), 71–7.

Aural comprehension deserves special attention because it is the easiest to test of the skills that can be trained in a language laboratory, because it is a valuable skill to acquire, and because it is directly related to the use in the laboratory of a variety of linguistic material for listening. The function of the language laboratory is to increase a learner's exposure to the foreign language.

In one basic type of listening compehension exercise transcription of the text is required, in another questions are set on the text. Partial transcripts are more valuable than complete ones. It is less useful to require students to answer questions from memory. The author describes a test developed in conjunction with a course in which comprehension exercises were prominent. Multiple-choice questions were used to test the recognition of the spoken version of familiar linguistic forms. The indeterminate true/false type of test was adopted, and the examination took place in a language laboratory. Multiple-choice questions enable the test as a whole to be used to evaluate its parts. [The author describes the construction and marking of the test in some detail, and examines the difficulty of setting a norm or pass-mark.]

42. **FARRINGTON, Brian**
A computer-marked examination in French prose composition for first-year university students *Modern Languages* (London), **53**, 2 (1972), 55–63.

The article describes an examination presented in the form of multiple-choice questions (MCQ) for first-year students at the University of Aberdeen. The type of MCQ selected is a variety of an indeterminate/true/false test of 100 items lasting two hours. The items consist of suggested translations of sections of a text in English which the candidate may accept, reject or leave unmarked, scoring +1 for a correct, −1 for a wrong answer, and 0 for a blank. The examination can be marked and the results analysed by computer. The validity of the test is considered by comparing the results with those obtained for the same groups of candidates from conventional examinations. It is shown statistically that for all but the most brilliant candidates the MCQ is at least as efficient as

traditional examination techniques in predicting later performance. Various incidental advantages of this type of test, and some objections to it, are also discussed. [Bibliography.]

43. **FARRINGTON, Brian**
What is knowing a language? Some considerations arising from an advanced level multiple-choice test in French *Modern Languages* (London), **56**, 1 (1975), 10–16.

The principles followed and techniques used in the devising of an objective test at advanced level are described. The test is used for first-year university students, consists of 90 items and takes two hours. It is marked by computer. Translation is avoided as much as possible. Lexical knowledge is measured by testing knowledge of the collocability of words or their paradigmatic relations [examples]. Syntax and word order are tested by items of rebarbative appearance, which test ability to recall the correct answer rather than merely recognising it. Negative oriented items formed the majority, the candidates having to select the one wrong option from among four acceptable ones, instead of the right form among four wrong. This means the candidates are faced with far fewer unacceptable pieces of French, and are openly encouraged to work by a process of elimination. [Mistakes previously made by comparable students are poor distractors – argument against the conventional five-choice item.]

The Discrimination Index was found to be low for language items compared to other subjects, probably because of the nature of language knowledge which is difficult to define, and most of which at an advanced level has been unconsciously assimilated rather than consciously learnt. Two types of linguistic proficiency at advanced level are distinguished: conscious acquisition, and the ability to generalise or recognise types. Some items failed by penalising the best (ie the latter) candidates. [Appendix gives correlation between scores in this test and the usual prose composition examination.]

44. **FELTKAMP, H W**
Kwantificatie van tekstbegrip [Quantification of comprehension skills] *Levende Talen* (The Hague), **304**, (1974), 34–50.

An analysis is made of German comprehension questions taken from official examination papers at late secondary level. Some weaknesses of multiple-choice comprehension questions are noted. Twelve specific proposals (seven aimed at the Central Institute for Test Development and five at teachers) are given.

45. **FOLLAND, David, and David Robertson**
Towards objectivity in group oral testing *English Language Teaching Journal* (London), **30**, 2 (1976), 156–67.

Group oral testing of the English of students studying business and

administration in Finland is described. The oral test is part of a wider test including composition and reading comprehension, and aims at a precise assessment of practical ability in spoken language. The group test is for a maximum of seven, who are tested in free discussion by two native English teachers; the stimulus is a taped report devised to cover a range of topics and with a suitable linguistic content. Arguments against oral testing are briefly answered and its advantages enumerated. [The marking system is described; adaptations of the test for other situations are possible.]

46. **FORREST, Ronald**
Objective examinations and the teaching of English *English Language Teaching Journal* (London), **29**, 3 (1975), 240–6.

The effects of introducing objective testing into English examinations for the West African School Certificate were examined by means of a questionnaire sent to teachers in Ghana, Nigeria and Ethiopia (in Ethiopia the School Leaving Certificate Examination is now entirely objective, but in Ghana and Nigeria the examination is only partly so). [Details of responses as to changes in teaching methods and teachers' attitudes as a result of the introduction of these tests.]

Although half the respondents considered that objective examinations had had a good effect on their teaching, the effects have been very mixed and were in many cases adverse (ie too much class time was spent on working at objective tests). More teacher training could help, as it was the better trained teachers who found that the new examinations gave them scope to improve their teaching.

47. **GIPPS, Caroline**, and **Elizabeth Ewen**
Scoring written work in English as a second language: the use of the T-unit *Educational Research* (Windsor, Berks), **16**, 2 (1974), 121–5.

For the purposes of scoring essays written in a second language, two of the most important considerations are the intelligibility and the structural complexity of the writing. Various disadvantages are inherent in the use of clauses and/or sentences as a basis for analysing structure in written work; a more satisfactory technique was developed by Hunt in America, using what he termed a 'minimal terminable unit' or 'T-unit' [details]. This technique was applied in the scoring of the NFER open-ended writing and speaking tests, which formed part of the battery of 'Tests of English Proficiency for Immigrant Children'. During the development of these tests, the battery was administered to Asian children, for whom it was found that the average T-unit length in writing and speech increased with increasing length of stay in Britain. (Average T-unit length has been found by Hunt and O'Donnell to increase with age for children writing and speaking in their native language.) The results of the NFER testing indicated that much of the development of proficiency in both the speech

and writing of the Asian children took place after three-and-a-half years in Britain. (These findings were essentially a by-product of test development, and therefore must be viewed with caution.)

48. GOMMES-JUDGE, Anne
Evaluating students' progress in language learning *Audio-Visual Language Journal* (London), **5**, 1 (1967), 21–5.

The principles of test CGM 62 (by Mialaret and Malandrin) can be applied to testing students' progress. The writer uses four tests, covering (a) oral and written comprehension, through translation into English and rapid dictation, (b) oral and written expression, through free composition and oral question and answer. Marks obtained are then plotted on a circular diagram so as to produce a quadrilateral, graphically representing over-all performance. [Examples of scoring procedures are given.]

49. HANDFORTH, D A
Conversation or catechism *Språk og Språkundervisning* (Oslo), **5**, 1 (1969), 11–14.

Not only is the practice of demanding a full sentence in answer to a question to be deplored, but also the kind of exercise which demands questions suitable to prompting replies as part of a small dialogue. Such testi g by asking for conversation-questions may also prove unsuitable as som pupils may well try to produce a dialogue reflecting English as a living language and not the pattern form set. The examiner is then faced with the classic dilemma of those who try to test more than one skill at a time. As some teachers are also doubtful of the value of filling-in tests, a suggested contextualised form is illustrated, giving a brief series of sentences forming a context first, followed by short prompt and introductory question words, from which pupils are asked to construct questions.

50. HARDIN, James N
A note on evaluation of student proficiency in the neglected languages *Foreign Language Annals* (New York), **4**, 2 (1970), 171–2.

An increasing number of American students are competent in non-European languages and some universities are accepting a knowledge of one of the lesser-known languages in fulfilment of their language prerequisite for advanced degrees. The Defense Language Institute has devised Defense Language Proficiency tests in over 200 languages.

The tests assess aural comprehension and reading comprehension and require only a technician who can play the tape-recordings, monitor the tests and grade them with a key. These particular tests are not available for civilian use, but similar tests are clearly needed. Cooperation between university language departments and institutions concerned with non-European languages might meet this need.

51. **HARRISON, Andrew**
Written production at the advanced level: testing an alternative to prose *Audio-Visual Language Journal* (Birmingham), **12**, 3 (1974/75), 179–83.

The prose as a test of written production in a foreign language should not lightly be abandoned. Alternatives, in the form of stimulus material or a battery of objective tests, are discussed but are felt to be less satisfactory. The prose is convenient for teachers and examiners and is an efficient testing device. An 'A' level candidate should be able to choose between different kinds of expression in writing, but must be restricted for the sake of a consistent mark scheme. An experimental test in French (called a 'reorientation' test) tried out in 1974 is described as a possible alternative to the prose. A passage was divided into two parts and candidates were asked to rewrite each part from a different point of view [details]. Language had to be suited to the particular account being written. Key-words, printed in italics in the text, represented 'events' in the story which the candidates were told to make the basis of their answers. About a third of the candidates included irrelevant material or overwrote, but nearly all understood the rubric. An ability to understand different points of view is vital in communication, hence this is felt to be a useful test. [Details of marking scheme.] [References.]

52. **HENNING, Grant H**
Measuring foreign language reading comprehension *Language Learning* (Ann Arbor, Mich), **25**, 1 (1975), 109–14.

Seven correlates of reading comprehension were examined for predictive validity, difficulty and discriminability. Twenty-seven Iranian college students were administered a series of eight test batteries so that there was a repeated measure sample size of 178. Cloze-recognition, synonym-antonym selection, multiple-choice sentence selection, question-written response and grammaticality of response predictors were examined at two levels of context difficulty. Intermediate difficulty synonym-antonym selection and multiple-choice sentence selection exhibited the highest correlation with the criterion.

53. **HILL, IL**
Multiple choice testing of the recognition skills in French *Babel* (Melbourne), **6**, 1 (1970), 12–16.

Reading and writing are still the most frequently tested foreign-language skills. The testing of recognition skills of reading and listening are illustrated here by objective multiple-choice methods. [Examples given.] Questions on reading passages are in English so that ability to express oneself in the foreign language is not tested at the same time as ability to understand. An anlysis is provided of the correct answers and distractors chosen by the students, to show the thought that has to be given to

providing suitable distractors. It is concluded that preparing and analysing such tests individually is too heavy a burden for any teacher and that a testing centre is needed from which items of guaranteed reliability could be obtained as needed.

For too long dictation has been the only means of testing listening ability and it tests knowledge of spelling and grammar as well. Listening comprehension lends itself to the multiple choice technique of testing. Pictures can be used, the student selecting a sentence which accurately describes the picture. The aural comprehension passage still has a place for the more advanced students when foreign language retention is also being tested.

54. **HILL, L A**
Objectives tests *Journal of the Teachers of English in Ethiopia* (Addis Ababa), **4**, (1968), 2–3.

As memorisation of model answers has been one of the ways in the past by which students attempted to pass examinations without adequate knowledge, and short answer tests can be too remote from real uses of languages, types of test are needed (and exercises too) which do not lend themselves to parroting but which show whether a candidate can use a language for practical purposes.

Suggestions are for essay-type tests in which the student is given detailed outlines for his composition and objective tests which are contextualised so that the candidate is faced with various situations to which he has to respond.

The author suggests an examination including (1) a test of aural comprehension, (2) an oral production test (response to stimuli), (3) a reading comprehension test, (4) written composition tests again responding to various stimuli which are, as far as possible, natural.

55. **HUTCHINGS, Geoffrey**
'Colourless green ideas': multiple-choice vocabulary tests *English Language Teaching* (London), **25**, 1 (1970), 68–71.

A sentence in a battery of multiple-choice test items has the test situation as its context. This includes such features as the examiner's intentions and the pupil's understanding of these. Imaginative candidates can find a context for almost any sentence. Where there are cultural differences between examiner and candidate, such tests are unreliable. The crux of the tester's problem is to make a context full enough to be unambiguous. The objectivity of multiple-choice tests is illusory because they lack a full context. Distractors sometimes provide clues. The less elementary the level, the harder it is to find plausible distractors. There is the danger of encouraging the idea that synonyms are interchangeable. Testing production vocabulary is more difficult than testing recognition vocabulary. One is

reduced to testing semantic collocation. Perhaps multiple-choice vocabulary tests inhibit experiment in the use of language.

56. **INGRAM, Elisabeth**
English standards for foreign students *University of Edinburgh Bulletin* (Edinburgh), **9**, 12 (May 1973), 4–5.

An investigation into the standard of English of students who are not native speakers of English, and their academic success, has been going on at Edinburgh University since 1968/9. All first-year students from abroad are required to take a test of English as a foreign language, though there is no question of passing or failing. The test used is the English Language Battery (ELBA), developed and administered by the Department of Linguistics. Part I tests aural comprehension; Part II is written and tests grammar, vocabulary and comprehension. Each testing session lasts two-and-a-half hours. [Details given, with a table of the number of students tested over four years, and the numbers for whom examination results are known.] Two other tables indicate clearly that the higher the score on ELBA the greater the chance of passing academic examinations. Students scoring 70 per cent and above in general do well. Almost half those who score below 50 per cent fail their final examinations.

57. **IRVINE, P, P Atai and J W Oller**
Cloze, dictation, and the Test of English as a Foreign Language *Language Learning* (Ann Arbor, Mich), **24**, 2 (1974), 245–52.

The Test of English as a Foreign Language (TOEFL) was taken by 159 non-native speakers of English in Tehran, Iran. They also took a cloze test and two dictations. The former was scored for both exact and contextually acceptable responses. A correlation of 0.94 between the two methods recommends the exact method for non-native teachers of ESL and concurs with the result of a study by Stubbs and Tucker (1974). Cloze scores correlated with the combined dications at 0.75. Cloze score by either method was more highly correlated with the TOEFL total score (0.79, acceptable word method; 0.78, exact word method) than was the combined dictation (0.69). Confirming earlier research by Darnell (1968) and Oller and Conrad (1971), both cloze and dictation correlated better with the listening comprehension than with any other subsection of the TOEFL. This tends to confirm the suspicion that the listening comprehension section, which is a highly integrative and pragmatic task, is more valid than other sections of the TOEFL. On the whole, listening comprehension, cloze and dictation were more highly correlated with each other and with the remaining part scores on the TOEFL than the latter were with each other. This last fact provides strong support for integrative testing procedures. [Tables; references.]

58. JAKOBOVITS, Leon A
A functional approach to the assessment of language skills *Journal of English as a Second Language* (New York), **4**, 2 (1969), 63–76.

Work in language testing to date has been oriented towards the assessment of linguistic competence rather than communicative competence. Linguists have generally maintained that the study of linguistic competence *per se* need not concern itself with social-psychological factors and the development of language tests as measures of linguistic skills derives its justification from this position. The adequacy of such a position is now challenged as a result of recent work in linguistics, particularly in semantics. Language 'codes', for instance, applicable to social situations have been recognised.

Three levels of meaning are here distinguished to account for the minimum range of linguistic pheonomena involved in communicative competence: linguistic, implicit and implicative meaning. A classification scheme is proposed for identifying the functional elements to be taken into account in the study of communicative competence, and some methodological approaches suggested for developing a testing program designed to assess skills involved in the use of language.

59. JAMES, Carl
Linguistic measures for error gravity *Audio-Visual Language Journal* (Birmingham), **12**, 1 (1974), 3–9.

If linguistic criteria could be used for the objective measurement of the language used by examination candidates exams would have definite advantages over so-called 'objective tests'. In fact, it is only in the scoring that tests can be called objective – their main drawback lies in defining the difficulty of each item. In exams, however, linguistic difficulty is irrelevant because the candidate can produce target language forms of any complexity he chooses. Another drawback of tests is that the pupils' expressiveness is stifled and what emerges is 'language-like behaviour'; exams, on the other hand, produce language in the context of the drive to communicate.

The test writer is concerned with homogeneity of difficulty, the exam marker with difficulty as reflected in the candidates' errors. A scale of degrees of errors is implicit in exam marking, whereas a test answer is either right or wrong. Linguistic considerations attendant on deriving error-weightings are discussed: (1) frequency; (2) consistency; (3) deviance, and (4) rule range. Under (3) it is suggested that the establishment of degrees of deviance would make measurement less subjective [reference to Chomsky's strict subcategorisation rules and selectional rules, which are analytic techniques which examiners could use] : using this process, a range of error gravity (4) can be established, more serious errors being those which entail further, dependent, errors.

Learners' utterances should not be matched with target-language forms and labelled deviant, but should be seen as a fluctuating, continuous

learning process. A new kind of contrastive analysis between transitional competence and native-speaker utterances would be helpful in preventing 'fossilisation' in L2 learning. [References.]

60. **JEANNERET, René**
Le test CGM 62 *Bulletin de la Commission Interuniversitaire Suisse de Linguistique Appliquée* (Neuchâtel), **2**, (1967), 19–28.

The test CGM 62 was devised for CREDIF (known by the initials of its compilers and the year of compilation) to examine the knowledge of beginners in French as a basis to drawing up homogeneous classes of students. [Details of the test are given.] The results give five levels of competence. At Neuchâtel some 120 students had taken the test for different purposes: classification on entering the commercial school; a final examination; a comparison between traditional and audio-visual teaching methods. The tests gave satisfactory results for the three purposes, in the third case showing more rapid progress in the audio-visual than in the traditional classes.

61. **JOHANSSON, Stig**
An evaluation of the noise test: a method for testing overall second language proficiency by perception under masking noise *IRAL* (Heidelberg), **11**, 2 (1973), 107–33.

The Department of English at Lund University, Sweden, has attempted to evaluate the method developed by Spolsky and others (1968) for testing overall second-language proficiency under masking noise. Such tests are possible because of the high degree of redundancy in natural languages. Certain alterations to the type of noise test and the scoring system were made, partly in order to adapt it for Swedish students of English, and a more consistent attempt was made to include sentences 'rich in redundancy'. Nevertheless, the technique of the two tests was still fundamentally the same, so that the Lund conclusions are applicable to the Spolsky or Indiana test. After a re-test only one month later it was found that there was not a good correlation between scores in the noise test and general English proficiency, particularly with low-score students. Writing speed, age and listening practice of students, as well as psychological factors, appeared to influence the results of the noise test, which is, in any case, possibly too artificial to correspond to aural comprehension ability in a natural language situation. A number of students who repeatedly failed in the noise test received high scores in a test consisting of dictation and traditional listening comprehension.

Lund University only recommend the noise test for certain categories of students, eg radio operators, interpreters and secretaries. They have found that dictation tests concentrating on specially selected words and expressions in recordings with no artificial distortion are a far superior method of testing overall second language proficiency. The usefulness in

most cases of a global test of second-language proficiency is questioned. [Correlation tables. Bibliography. Sample of the Lund noise test.]

62. KAMRATOWSKI, Joachim
Leistungstests im Sprachunterricht [Achievement tests in language teaching] *Englisch* (Berlin), **6**, 4 (1971), 97–101.

Achievement testing is now accepted but criticism is still levelled against the methods of test compilation and administration. In present-day education such tests are becoming more and more an instrument for self-criticism. The objectivity and limited scope of a test are its strongest points. It must be reliable and must correspond to certain recognised norms which give it general validity. It must be subjected also to careful evaluation. In addition to tests which may be applied to, for example, all pupils in their fifth year in one district, there is a place for informal classroom tests on active vocabulary, passive vocabulary, aural comprehension, reading comprehension, grammatical structures and spelling. There are many types but two are suitable for foreign languages: multiple-choice tests with four choices [advantages and disadvantages discussed] and gap-filling tests which can be used to check active vocabulary and spelling. [Several examples are given of test forms which have been validated and should encourage a beginner in test construction to develop tests for himself and try them out in the course of his teaching.]

63. KANKASHIAN, Ibrahim
Objective vocabulary testing *TEFL: Bulletin for the Teaching of English as a Foreign Language* (Beirut), **1**, 4 (1967), 1–2 and 4–5.

Two types of objective vocabulary test are illustrated; a 'true or false' and a multiple-choice test. By examples the author shows what constitutes effective and poor tests and recommends a system of marking to eliminate wild guessing. In 'true or false' tests there may be unintentional difficulties if too many words are used indicating indefiniteness in time, degree, frequency, extent, size and amount, as the student's estimate of such abstractions may be quite different from the teacher's. In multiple-choice tests homonyms and pairs of opposites are to be avoided. All distractors should be plausible. The correct answer should not be longer or shorter than the other words.

64. KOBERSKY, Eva
Testing recognition and production with oral stimuli (experiments being carried out at the Department of Linguistics at the Centre Universitaire in Mons, Belgium) *Contact* (Aarau), 12, (1969), 22-4.

Potential students at the Ecole d'Interprètes Internationaux are tested for their knowledge of French. They are also given an aptitude test for language learning, as they are required to learn two completely new

languages to interpreter's standard in two years. The aptitude test, which has to be suitable for students of various mother tongues, aims to reveal ability to recognise and reproduce foreign phonemic contrasts. Hungarian was chosen for the test. Through three batteries of five minimal-pair sentences, recognition and reproduction of intonation, double versus single consonant, and long versus short vowel in unstressed syllables, are tested. Instructions for the test are given in French and it lasts four minutes. Native Hungarian speakers score the results. All students are tested at the same time in a tele-directed language laboratory. The aptitude test is validated at intervals with revised tests containing new samples. Comparison of results should show if the aptitude test has successfully predicted ability to study foreign languages.

65. **KOLLÁRIK, Teodor**
Niektoré poznatky so zist'ovaním spôsobilosti učit' sa cudzie jazyky [Some comments on aptitude testing for foreign-language learning] *Cizí jazyky ve škole* (Prague), **16, 5** (1972/3), 206–11.

An aptitude test, constructed on the basis of Carroll's MLAT, was administered to 113 adult students (aged twenty-nine to forty-nine) at the Institute for the further training of managerial staff in Slovchémie, Bratislava, Slovakia, at the beginning of their part-time release foreign-language course. The results of this test were matched with the result of the final examinations of the course (both oral and written) and fully proved the predictive validity of the aptitude test. It was also investigated to what extent the results of the aptitude test and of the examination results could be compared with the level of intelligence. IQ was measured by means of a non-verbal test which was not considered fully adequate, so its results were only of an informative and indicative value. The group of students with a high IQ achieved better results in the test and in the examination than the group with a lower IQ. The comparative results, however, do not indicate whether the good results in the examination are primarily due to a high level of IQ and/or to an aptitude for learning languages (apart from other factors determining success in learning). But the relations existing between the parameters, which have been computed, justify tests for assessing the aptitude for learning languages and demonstrate their potential.

66. **KOLLMANNOVÁ, Ludmila**
Test slovní zásoby [Vocabulary test] *Cizí jazyky ve škole* (Prague), **16, 9** (1972/3), 385–93.

In the design of English vocabulary tests a prime factor is congruence of objectives and content. Lexical items should be selected according to fixed criteria of frequency, difficulty and degree of interference from the native language in a given context. The choice of context is dictated by the student's knowledge of grammar, and the lexical content should be drawn from real situations. The choice of test item depends on the target student, the aims, and the criteria of objective testing.

67. **KOMOROWSKA, Hanna**
Zasady konstrukcji testów a poziom zaawansowania językowego vczniów (klasy II, III, IV) [Principles of test construction for pupils at advanced level in languages. Classes II, III, IV] *Języki obce w szkole* (Warsaw), **18**, 2 (1974), 70–80.

A battery of tests suitable for advanced classes in foreign languages in a General Education Lycée is described. For Class II these include: phonological, lexical, morphosyntactic, communication, speech, aural and written comprehension, and writing skills tests. For Classes III and IV: similar types but with different components and adjusted to suit these levels. [Examples of each type of test.]

68. **KÜNNE, Wulf**
Erstellung eines fremdsprachlichen Aussprachetests: Probleme und Methoden [Devising a pronunciation test for a foreign language – problems and methods] *Die Neueren Sprachen* (Frankfurt am Main), **22**, 7 (1973), 372–80.

The importance attached to oral work is not matched by ways of measuring it objectively. One aspect of oral use, pronunciation, can be tested objectively. A test for first year teacher-training students studying English is described; it is diagnostic and the aim is to achieve the pronunciation of a native speaker. The test comprises dialogue passages for reading aloud, taken from MacCarthy's *English conversation reader, the* vocabulary being that of current usage. The test items are restricted to the main problem sounds; of the 11 items, six are pairs which cause difficulty; each item occurs between ten and 20 times in the test. [The test marker's version is described and illustrated. Bibliography.]

69. **LASCH, Miguel**
A prognostic test of language-learning ability *English Teaching Forum* (Washington, DC), **5**, 3 (1967), 10–14.

The aim was to find a means of selecting teacher trainees (in Chile) on a basis of potential rather than of achievement. The skills selected for testing were aural discrimination, oral production, grasping semantic distinctions, and grasping and manipulating structural devices. The stimuli were recorded, but not the students' answers. [The author gives details of eight different types of tests: repetition of a series of facts, repetition of words containing unfamiliar sequences of familiar sounds, pronunciation of tongue-twisters, imitation of vernacular mispronunciations, repetition of nonsense words, imitation of rhythms, imitation of musical tunes, and answering of questions in English.] A written recognition test was given as an extra. Listening and speaking were tested as a single skill. The students' ability to recall after drilling was not tested. Two or three evaluators listened to the candidates' performances.

The candidates with the lowest scores failed the course. The test can

therefore predict failure. Almost all who got an average of 65 per cent on the tests succeeded in the first year of study.

Additional sections will be included in future tests. [Detail.]

70. **LEUTENEGGER, Ralph R, Theodore H Mueller,** and **Irving R Wershow**
Auditory factors in foreign language acquisition *Modern Language Journal* (St Louis, Missouri), **49**, 1 (1965), 22–31.

A study to measure the auditory factors active in adult English-speaking beginners learning foreign languages was made. Seashore Measures of Musical Talents were employed in conjunction with intelligence of aptitude tests on 283 students of French and 177 students of Spanish at the University of Florida. Multiple regression techniques were applied to the data obtained from the nine subtests. Fifteen variables were selected of which only Tonal Memory and Total Reading Score were found significant in predicting foreign-language acquisition, measured against the criteria of students' average scores in tests of achievement which formed part of their language courses.

71. **LEVINE, Josie**
An outline proposal for testing communicative competence *English Language Teaching Journal* (London), **30**, 2 (1976), 128–35.

A test is proposed for middle-school children in multiracial classes in Britain, to assess their proficiency in using English for purposes of learning, thinking and communicating, and thus enable teachers to plan their future teaching; the model should also be applicable to tests of communicative competence in FL and SL situations. Language tests generally only test a small number of the combinations which make up the concept of full language skill; the required model would test not only a student's knowledge of the language and its contextual restraints but also the degree to which he can communicate effectively in specific circumstances. Objective marking against a grid is recommended, to reveal the degree of proficiency in a defined setting, with further discrete-item tests to specify in more detail students' lack of knowledge in any of the component language-skill areas, ie discrete-item tests backing up context tests. [Brief discussion of the test battery, including a rating sheet and possible test procedure.]

72. **LÖSCHMANN, Martin,** and **Marianne Löschmann**
Kontrollverfahren und -formen für das stille Lesen [Testing procedures and test forms for evaluating silent reading] *Deutsch als Fremdsprache* (Leipzig), **12**, 4 (1975), 216–22.

Silent reading can only be tested through textual understanding. Such tests should encourage the student to do his reading more effectively. Results from tests involving translating, speaking and writing are only partly relevant to evaluating silent reading. The degree of difficulty of the tests

has to be increased progressively to keep pace with more complex texts. Depending on the level of proficiency, tests can be performed in class or at home. Extra-linguistic tests elicit non-linguistic reactions [examples] and tend to concentrate on the wider context rather than on details. For advanced learners a combination of linguistic and extra-linguistic tests is recommended. In the case of receptive tests, such as multiple-choice questions, the student has to decide if a given utterance agrees with the text he has read; the test thus acts at the same time as another exercise. [Disadvantages listed and modifications suggested.] Reproductive tests require the student to repeat parts of the text in answer to questions, which can be a superficial and formal exercise. Productive tests are meant to elicit independent, even divergent utterances, and represent the most demanding form of test, offering a wide range of procedures [some listed]. [Some other less common testing procedures are briefly described and the use of the mother tongue is discussed.]

73. **LOWE, J Edgar**
Are you fair in your grading? *English Teaching Forum* (Washington, DC) **10**, 2 (1972), 9–14.

Testing and evaluation procedures in the teaching of English as a foreign language need to be improved. Grades should not be given to encourage or stimulate students. Tests should provide the teacher with information he cannot get otherwise. Improper weighting of scores is a source of unfairness, and there should be coordination among English teachers in the same school. [The author considers the advantages and disadvantages of oral/aural, essay, and objective tests and examines in detail two different ways of testing vocabulary.]

In deciding which language skills to test, three things should be kept in mind: the material taught in class, the record of progress made and problems experienced and the record of recurring errors. [He describes the construction of a test under the headings: Step 1 Planning the test; Step 2 Writing the items; Step 3 Integration.]

Effective testing is a stimulus to study. It encourages self-examination and better preparation, and tends to build up the student's confidence in the teacher. It reveals what the students have learned, tells the teacher about the effectiveness of his methods, gives a measure of individual differences, and enables the teacher to evaluate a student better. [The author concludes by listing the requirements for a good test, and by enumerating good testing procedures.]

74. **McLEOD, John**
Uncertainty reduction in different languages through reading comprehension *Journal of Psycholinguistic Research* (New York), **4**, 4 (1975), 343–55.

In an explanatory study into the feasibility of constructing a multilingual

reading comprehension test, parallel cloze tests were constructed in Czech, English, French, German and Polish. Groups of children between the ages of eight and fourteen were tested on the version in their own mother tongue, in Czechoslovakia, Canada, France, Germany and Poland. Test performance was assessed by means of a method which estimates the child's 'contribution' to overall language redundancy. Results suggest that the method is both reliable and valid and has potential for the assessment of bilinguality of reading comprehension.

75. MADSON, Harold S

New alternatives in EFL exams, or 'How to avoid selling English short' *English Language Teaching Journal* (London), **30**, 2 (1976), 135–44.

As it is not yet possible to do away with the school leaving examinations in Africa, the most satisfactory alternative to existing examinations is felt to be one which combines the thorough, systematic evaluation of reading comprehension, structure and usage (a condensed version of one of the present examinations) with productive, integrative evaluation. The combination of integrative and discrete-point tests provides the broad scope and analytical features of the older all-objective test as well as a natural synthesising of language. The productive element should encourage communicative activities to return to the classroom. [Discussion of test content; references.]

76. MAMAN, André

Avantages et limites des tests objectifs dans l'apprentissage du français, langue étrangère [Advantages and limitations of objective tests in the early stages of French as a foreign language] *FIPF* (Sèvres), **6/7**, (1973), 82–99.

Objective French tests are prepared by teams of specialists to provide standardised mark patterns for assessing either the aptitudes for French or attainments in French of student populations in one country (here, the USA). They can only be used on a large scale and must be valid, reliable, scorable, inexpensive and easy to administer. The tests must be used for the purpose for which they are designed and not casually. [Other criteria for their use are defined.] Thus they should provide prompt identification of inadequate progress and of serious learning difficulties, as well as an accurate guide as to the suitability of applicants for academic courses. The administration should redress injustices resulting from the clumsy use of objective tests. Teachers' and students' criticisms of the tests are given, and possible disadvantages for indigent students and some ethnic and social subcultures are outlined. National tests cannot be 'exported' for use by speakers of a different mother tongue. It is desirable that different tests be used for candidates who use a particular learning mode, such as home study.

77. **MATTHEWS-BRESKY, R J H**
Translation as a testing device *English Language Teaching* (London), **27**, 1 (1972), 58–65.

Translation from the mother tongue into English is useful as a device for testing linguistic items, although there is little to be said for it as a means of teaching English. [Four sets of examples given for testing Germans in English: concepts which are plural in German but singular in English, and vice versa; English word-order and especially the position of adverbs; selection of the present continuous and the present simple; and the omission of commas in English.]

Though target-language testing has the advantage of working totally in the language that is tested, the format of a poor test can actually promote interference. Moreover, target-language questions involve unnatural language. Often it is too obvious to the pupil which language item is being tested. The purpose of a test is not to find out whether the pupil can practise a grammatical point when called upon to do so, but to discover whether he incorporates the point into what he already knows of the language. In a translation test the instructions are unambiguous, there is no temptation to make 'unnatural' mistakes, and the pupil does not know what is being tested. [Examples are given both of target-language tests and of tests by translation.] The latter are not tests of translation itself.

Testing by translation works only if the pupils are literate in their mother tongue and if there is a tradition of approximate linguistic equivalence. Further conditions are that it is not comprehension which is being tested, that the incidental features of the material to be translated must not be difficult relative to the class standard, that on the relevant point one response only is acceptable, and that between the mother tongue and the target language there are points of grammatical divergence best tested by controlled translation.

The level of extraneous difficulty should be kept low and loopholes should not be left by means of which the pupil can avoid the issue. Not all features of English structure can be effectively tested by translation. Translation exercises can be made more interesting by placing them in the context of the interpreter's situation. [Example of German policeman interrogating English criminal.]

78. **MOHAN, Bernard A**
Comprehension testing as semantics plus induction *Linguistics* (The Hague), **115** (1973), 93–104.

The author is concerned with establishing techniques for demonstrating sentence comprehension that do not rely on non-linguistic responses. Techniques need to be measured against an ideal of understanding all the grammatical relations in a sentence and the semantic reading assigned to sentence by a semantic theory. Furthermore a method must provide arguments that are inductively strong [examples and discussion]. The

semantic theory assumed is that common to Leech, *Towards a semantic description of English*, and Katz, *Semantic theory*. The test evolved requires children to operate criteria of 'implication' and 'inconsistency' in the sorting and matching of correct and incorrect paraphrases to assertions contained in a passage of prose. Examples are given to show the intimate relations between the method of eliminative induction required to give logical rigour to comprehension testing and the theory of semantic structure advanced to account for the process of comprehension. It is argued that tests which ignore semantic theory are not truly concerned with comprehension and that those which ignore inductive schemata are not truly tests.

79. **MÖHLE, Dorothea**
Der Einsatz informeller Tests im Französischunterricht [The introduction of informal tests in the teaching of French] *Die Neueren Sprachen* (Frankfurt), **24**, 1 (1975), 2–18.

It is recognised that some tests are not simply a method of grading pupils, but a positive part of the teaching process. For reasons of time and objectivity, most tests in an orally based course must be written ones. The French language presents special problems because of the abundance of its grammatical forms, even at an elementary level, the interdependence of form and context and the very different systems of the oral and written languages. It is therefore important that the tests devised should cover specific points, rather than make a global assessment of failure or success. A series of tests of verb, article and adjective agreement is given as an example. In these tests the pupil is asked for productive answers of a strictly controlled nature, ranging from inserting one word in a gap to supplying complete statements or questions. As the degree of difficulty increases, the possible answers are too varied to be marked with absolute objectivity, but at all stages it is valuable to isolate specific elements of free utterance and assess the pupil's proficiency in handling them. [Examples.] [References.]

80. **MOLLER, Alan**
Validity in proficiency testing *ELT Documents* (London), **3**, (1975), 5–18.

Proficiency is concerned with displaying the ability to use language in a variety of situations. Proficiency testing is a form of attainment testing together with achievement and diagnostic testing. [Distinction between achievement, diagnostic and proficiency tests, and between norm- and criterion-referenced tests.] Five types of validity are discussed: construct, content, concurrent, predictive and face validity. Validity criteria are applied to two measures used extensively overseas, the subjective assessment procedure and the English Proficiency Test Battery ('Davies Test'). [Description of criteria, samples of overseas students in Britain and tutors'

ratings of language ability. A sub-sample was tested for communicative proficiency.] The relationship between predeparture assessments and tutors' ratings after students had begun their studies is analysed [details]. It is concluded that the Davies Test was successful in 86 per cent of its predictions of language ability, and that subjective assessment is generally valid when administered rigorously.

81. **MOST, R L**
Testing aural comprehension *Rapport d'Activités de l'Institut de Phonétique* (Brussels), **6**, (December 1972), 42–59.

A study was undertaken of aural comprehension with specific reference to testing non-specialist students on completion of an intensive revision course based on audio-aural methods in English language at an elementary/intermediate level. The students were in the second year of their course in the Faculty of Economics and at the École de Commerce of the Université Libre de Bruxelles. The project covered the three academic sessions 1969/72. [Aims and objectives of the project are outlined and a detailed description given of the early tests prepared and administered during the first year, and the modifications that were required.]

In the new tests open-ended questions were replaced by a multiple-choice format and the load on short-term memory was substantially reduced. The test lasted a strict thirty minutes. In pretest trials, scoring was based on an unweighted system with two marks for each of the ten questions. Scores were concentrated towards the top end of the scale, which is not unusual in a proficiency/achievement test. In the sessional tests the results also showed high scores, in part because the passage chosen facilitated decoding by its logical arrangement and clear division into key points. In the later tests, items were weighted according to difficulty.

One of the aims of the project, the construction of a valid, reliable test, is nearing fulfilment, though performance in listening comprehension does not correlate highly with performance in tests of other linguistic skills. Further experiment is necessary on equivalent reliability (on whether tests are in fact of equal difficulty), and on the kind of response open to the candidate (a précis in the mother tongue might replace the multiple-choice format). It must be accepted that comprehension of connected discourse requires non-linguistic skills, in particular, abstract reasoning ability.

82. **NEHLS, Dietrich**
Report on the meeting of experts on language testing in adult education held at the residential college for adult education at Göhrde, W Germany, June 12–15, 1968 *IRAL* (Heidelberg), **7**, 2 (1969), 133–48.

This conference, organised by Nowacek, head of the federal testing centre in foreign languages of the German Adult Education Federation, provides

in its report a valuable summary of available aptitude and proficiency tests, showing contemporary aims and procedures. The conference was divided into two groups, one concerned with the testing of English as a foreign language and the other with testing French.

83. **NIŻEGORODCEW, Anna**
Przydatność tzw. testu prognostycznego jako metody pomiaru zdolności do uczenia się języka obcego u młodzieży licealnej [The usefulness of so-called predictive tests as a means of measuring the ability to learn a foreign language among secondary school pupils] *Języki obce w szkole* (Warsaw), 19, 5 (1975), 299–303.

A test (Test P) based on language aptitude tests already in existence was contrived to verify the pupils' inductive ability. The pupils had to apply the grammatical rules in their native language to three micro-languages [examples]. The test was given to a group of 89 pupils in the third and fourth class of a Polish secondary school as well as 78 in class I. The test was then retaken after six months. The correlation between term results in Polish, mathematics, history and chemistry and the results of Test P was higher than the correlation of Test P with the results in English. Test P did not therefore prove a very reliable means of predicting future ability in the learning of a foreign language. Further research is being carried out to discover the effectiveness of the test used at a more advanced stage of language learning.

84. **NOVÁK, Zdeněk**
K vývoji testu predikce výkonu v učení se cizím jazykům [The development of a prediction test of performance in foreign language learning] *Cizí jazyky ve škole* (Prague), 17, 1 (1973/4), 18–22, and 2, 66–74.

The author discusses some theoretical problems involved in the design of an aptitude test for the study of a foreign language by Czech children at the age of eight (before entering the third form of a Basic Nine-Year School with an extended foreign-language programme – Russian, French or German). The test battery has been constructed as a 'pencil and paper' type, following the structures of the Pimsleur Language Aptitude Battery and the Carroll Elementary Modern Language Aptitude Test, both of which have been designed for learners of the age of 11 upwards. The main problem was how to adapt the single items of these test batteries as well as the whole tests to children of a considerably younger age-group. The author describes in detail the final version of the test, in particular the single subtests, grouped into two sections of three. The experimental natural language chosen for the purpose of the test was Albanian because it is an Indo-European language and therefore belongs to the same family of languages, one of which Czech pupils are expected to learn; also because Albanian is considerably different from any of the languages which are commonly known in the area. The test has been administered to 168

children of the ages 8.0–8.8 at the beginning of their third year of school in September. The main statistical indicators are critically evaluated as to mean, standard deviation, dispersion of raw scores, coefficients of reliability and validity, and in the light of correlations especially between the test and school achievements in the native language (Czech) and in one of the foreign languages (French or Russian or German), according to the highly subjective school reports at the end of the winter semester in February of the following year.

85. **NÜBOLD, Peter**
Sprechfertigkeitstest: Lernzielgerechte Leistungskontrolle im Sprachlabor und ihre technologischen Probleme [Oral productive ability: goal-oriented learning in the language laboratory and its technical problems] *System* (Linköping), **2**, 3 (1974), 10–20.

Oral productive ability is accepted to be the major teaching aim of foreign-language instruction at the level of *Sekundarstufe I* in West German schools. It is usually by means of written tests, however, that a pupil's progress is judged. The reason for this contradiction is not so much a lack of oral production tests as the fact that such tests can be carried out adequately only in an AAC language laboratory and that the expense of carrying them out is enormous.

The application of a new recording device would reduce this expense to a reasonable amount. As the oral production of a whole test-group could be directly recorded onto a single magnetic tape or disc, production tests could be carried out in an AA lab as well. Scoring could be done anywhere at any time without the need to remove the tapes from the students' booths or to copy them tediously one after the other in order not to obstruct the lab. An additional advantage would be a reduction in the space required for the necessary filing of tests.

86. **OLLER, John W**
Dictation as a device for testing foreign-language proficiency *English Language Teaching* (London), **25**, 3 (1971), 254–9.

Evaluation of an ESL placement examination revealed that dictation scores correlated more highly with those of other parts of the test than did any other part. Dictation seemed to be the best single measure of the English-language skills tested. It provided a more comprehensive sampling of the integrative skills involved in understanding complex English structures than the more isolative and analytical objective tests.

Both Bloomfieldian and Chomskyan linguists have treated language as if it were a self-contained system. If language cannot be explained apart from its use as a medium of communication, analytical tests which remove linguistic units from the meaningful contexts in which they occur are less valid than integrative tests, which are more relevant to communication

skills. In dictation, a student is tested for his ability to discriminate phonological units, to decide on word boundaries, and to translate his analysis into a graphemic representation. [Bibliography.]

87. **OLLER, John W**
Cloze tests of second language proficiency and what they measure *Language Learning* (Ann Arbor, Mich), **23**, 1 (1973), 105–18.

Cloze tests have been used for many different purposes. They are good indicators of reading comprehension and of the readability of prose. But until recently almost all the research on cloze tests used native speakers of a language and almost all the studies were in English. Among problems not investigated with non-native speakers are those of the distance between blanks and of the grammatical function of deleted words. Multiple word deletion and the mode of presentation have been insufficiently studied with both native and non-native speakers. With non-native speakers the scoring method which counts only exact word-replacements as correct is inferior to a method which allows contextually acceptable responses. Similar information is obtained from cloze tests of various levels of difficulty; cloze procedure can be used to diagnose grammatical problems; translating a cloze passage from one language to another produces tests of equal difficulty for native speakers of the two languages and the responses of non-native speakers differ in kind as well as degree from those of native speakers.

Cloze tests of second-language proficiency are internally consistent, reliable, valid, and easily constructed, administered and standardised. Inter-scorer reliability needs to be investigated. Language skill cannot be separated from knowledge. Cloze tests measure a grammar of expectancy. In taking such a test a student makes an analysis of the information provided and synthesises a greater whole, which then helps with further analysis and synthesis. This is also what happens in speaking or writing. Cloze tests of second-language proficiency correlate best with measures of listening comprehension. The views of Upshur and Spolsky are compared on the effect of the time factor on verbal processing. Any kind of language processing imposes time constraints. Memory constraints are one aspect of second-language speaking competence.

88. **OLLER, John W**
Pragmatic language testing *Language Sciences* (Bloomington, Ind), **28**, (1973), 7–12.

A distinction must be drawn between relationship messages and content messages [examples] . Analysis of linguistic communication has concentrated on content coding, but relationship messages may be more important in communication. The concept of a grammar of expectancy is basic to both types: expectancies play an essential role in the perception and production of language. Sequentially connected sentences are more easily

perceived and produced than isolated sentences. Predictability, to which creativity and surprise value are related, makes grammars of expectancy possible: the mind analyses its experience and learns to rearrange the components informatively. Linguistic creativity is more than the capacity to generate an indefinite number of mental objects (eg sentences), but entails the capacity to modify the generative mechanism. Intelligence, meaning and language result from creative interaction between an individual and his environment. Sentences and other linguistic units in relation to extralinguistic contexts have characteristics that sentences independent of these contexts do not have.

Discrete-point testing violates the principle that the whole is greater than the sum of its parts. If different subtests intercorrelate well, they are highly valid. Discrete-point tests are inferior to integrative tests, which require activities similar to those of normal communication [dictation and cloze tests are briefly discussed]. Integrative tests provide information on the student's deficiencies, are claimed to be easy to construct and administer, have clear application to teaching, and discriminate between native and non-native speakers better than discrete-point tests. [References.]

89. **OLLER, John W,** and **Christine A Conrad**
The cloze technique and ESL proficiency *Language Learning* (Michigan), **21**, 2 (1971), 183–95.

The cloze procedure is a test which measures integrative language skills. It has the advantage of requiring the student to perform tasks similar to those of native speakers in sending and receiving messages. It may be applied to teaching and is easy to construct and score. [Research with native and non-native speakers is briefly summarised.]

A cloze test was given to two groups of native speakers and various groups of non-native speakers of differing levels of proficiency in English. Results showed that when the exact-word scoring method was used, the test failed to discriminate adequately between native and non-native speakers, while it over-emphasised the differences among native-language speakers (in this case, freshmen and graduates). When any contextually acceptable word was counted as correct, the differentiation of non-native-language speakers was better, and there was closer correlation with external validating criteria. The ability to carry out language tasks may vary more widely among native-language speakers than it does among native-language and select non-native-language speakers. Thus specific ESL training for fairly advanced ESL speakers may be less effective than instruction in subject-matter of the type normally provided to native-language speakers. On the cloze test used, some of the advanced learners did better than some of the native-language freshmen. Cloze test scores correlate well with reading comprehension and dictation scores. [Bibliography and specimen cloze test passage.]

90. **OLLER, John W, and Virginia Streiff**
Dictation: a test of grammar-based expectancies *English Language Teaching Journal* (London), **30**, 1 (1975), 25–36.

A re-evaluation of data given in an earlier paper [see entry 86] with additional information as to test rationale, administration, scoring and interpretation. A theoretical explanation is offered to account for the usefulness of dictation as a measure of language proficiency; it is seen as a device which measures the efficiency of grammar-based expectancies by means of a two-stage process of synthesis and analysis. Since dictation activates the learner's internationalised grammar of expectancy, the central component of his language competence, it is bound to yield useful information on overall proficiency. [References.]

91. **PAGE, B W**
Another look at examinations *Audio-Visual Language Journal* (Manchester), **11**, 2 (1973), 127–30.

A serious drawback of the present examination system (particularly 'O' and 'A' level language examinations) is that boards are vague in their requirements and teachers must infer these from past papers. Examiners are criticised for having unfair expectations and ill-defined criteria for their criticisms. Present examination methods are not professional enough and the system favours those who have an inside knowledge of it.

How might an official syllabus for language examinations be set up? It is important to define the requirements for a minimum pass. A possible alternative would be a system of graded examinations at several levels which would enable even learners with a relatively low level of proficiency to obtain some recognition for what they had learned [further advantages are outlined].

92. **PAUELS, Wolfgang**
Englische Rechtschreibtests in Multiple-Choice-Form in Hauptschule und Gymnasium [Multiple choice English spelling tests in the 'Hauptschule' and the 'Gymnasium'] *Praxis* (Dortmund), **22**, 1 (1975), 34–41.

Multiple-choice spelling tests may, by presenting pupils with wrong spellings, interfere with the learning process. Moreover, their effect is greater with lower-ability pupils *(Hauptschüler)* than with higher-ability *(Gymnasiasten)*. To test these theories, two groups of 200 pupils in first year secondary were selected, one from each type of school. Each group was presented with the same series of four tests: in test 1 eight English words were tested by dictation; test 2 and 3 tested recognition of the same words by multiple choice answers; in test 4 the same words were to be inserted in the correct gaps in a short test. In tests 2 and 3 the *Gymnasiasten* showed a greater immediate improvement over their own score in test 1 than did the *Hauptschüler*, although only the latter continued to improve in test 3. This may indicate that the *Hauptschüler* take longer to react to visual simuli, but that the effect continues longer. In test

4 the *Gymnasiasten* made fewer extra errors, compared to their score in the two recognition tests, than the *Hauptschüler*, which seems to indicate that the latter were indeed adversely affected by the wrong spellings. On the other hand, the *Hauptschüler* showed a considerable improvement in test 4 over their score in test 1, showing that the visual reinforcement in tests 2 and 3, even allowing for the wrong spellings, still had a considerable effect. This point has relevance to the teaching of spelling to lower-ability groups. [Transcript of tests.]

93. PENDLEBURY, A C
Testing vocational English *English Language Teaching* (London), **24**, 3 (1970), 254–260.

The term 'vocational English' refers to the teaching of specific vocabulary connected with certain occupations together with basic English patterns and structures. The testing of this aspect of English should be integrated into a language-training course. Only one aspect of any area of learning should be measured by each question. Tests should be suitable for objective scoring and should be easily administered. [The author discusses listening, speaking, reading, and writing tests of the student's knowledge of the names of things, of the meanings of words describing activities or events, of the characteristics of objects, and of the meanings of words.] There is at present a lack of originality in tests used in language teaching. Testing can be fun: a variety of techniques is available.

94. PFAFF, Harald
Die Entwicklung eines Sprachtests im Rahmen einer linguistischen Theorie [The development of a linguistic test within a lingutistic theory] *Linguistische Berichte* (Braunschweig, Germany), **35**, (1975), 71–82.

The test should judge the receptive linguistic competence of children according to an ideal norm. The different performances of the candidates to the norm can be measured and tested using multiple-choice questions. The syntactic or semantic acceptability of sentences can be tested in this way [examples] . The choice of which of three or more sentences are formally or semantically similar can test a knowledge of transformations [examples] . Ability to describe ambiguity in sentences, to form words and to form subordinate sentences can also be tested [examples] . The test asking candidates to mark those sentences which have the same form showed no significant correlation with the candidates' IQ values, but only nine items were used in the correlation. The test 'formal similarity' can, therefore, measure the receptive competence of some candidates.

95. PICKETT, G D
A comparison of translation and blank-filling as testing techniques *English Language Teaching* (London), **23**, 1 (1968), 21–6.

The author describes an attempt to discover whether blank-filling or

translation is the better means of testing linguistic ability. He deals with possible objections to the standard he used – the average mark per pupil gained over a half-year for classroom work. The procedure involved the construction of two tests, the identical answers to which were obtainable in one test by blank-filling and in the other by translation. The subjects were French-speaking Africans. Scores for the translation test were much nearer the half-year result than those for the blank-filling test.

Blank-filling cuts down the risk of making mistakes and tends to cover up the differences between the best and the worst examinees. Much time and paper would be needed to test all the features of the written language by blank-filling.

96. **PINEL, Nicole**
Test d'évaluation de niveau oral pour les jeunes débutants en francais [A test to evaluate oral ability for young beginners in French] *Français dans le Monde* (Paris), **79**, (1971), 14–19.

The *Lycée Français* receives children of mixed nationalities, predominantly French and English at various stages of ability in comprehension of and expression in French. A test of oral ability was therefore devised in order to classify the seven-year-old entrants into homogeneous groups. A basis of phonetic, lexical and structural knowledge was established without which a child could not enter the school. Coloured pictures were used and efforts were made to conduct the test as if it were a game so that the young children should not be upset by the test atmosphere. [Examples given.] It was early to draw conclusions from such a test but it clearly separated out children of widely differing linguistic ability. Those whose linguistic abilities were average were sorted by the test results into a traditional and an experimental class. The margin of error was subsequently found to be small.

97. **PLAISTER, Theodore H**
Testing aural comprehension: a culture fair approach *TESOL Quarterly* (Indiana), **1**, 3 (1967), 17–19.

Testing aural comprehension by pictures is unreliable as the student may meet with cultural difficulties and the test will be one of picture interpretation rather than aural comprehension. Tests in which students listen to a piece of prose and make notes before answering questions are misleading as this does not reflect the normal lecture situation.

It is important that any test should measure comprehension of a wide range of fluently spoken English rather than the understanding of specific points. A multiple-choice test has been devised whereby students select one of four diagrams to show they have understood a spoken lead. This test is efficient and measures only the students' linguistic ability. It is also easy to produce. [An example of the test is included in the article.]

98. **POLITZER, Robert L**
Towards a practice-centred program for the training and evaluation of foreign language teachers *Modern Language Journal* (St Louis, Missouri), **50**, 5 (1966), 251–5.

The MLA Foreign Language Proficiency Tests for Teachers and Advanced Students are now widely accepted as the norms according to which the preparation of the foreign-language teacher is to be judged. However, additional tests appear to be needed which measure the teacher's performance in the classroom (which to some extent can be done objectively in terms of the achievement of his pupils). Such tests would help to develop practice-centred training. At Stanford University micro-teaching procedures provide instruction in a specific classroom teaching skill to the trainee, who then applies it in a five- or ten-minute lesson to a small group of pupils. This technique can be applied to the retraining of language teachers, using as a basis the pattern of skills suggested by the MLA tests. Examples are given of how this can be done, particularly with the use of video-tape recordings. A syllabus is being developed for training teachers of French through the use of twenty-five micro-lessons of five to ten minutes each.

99. **POLITZER, Robert L**
Developmental sentence scoring as a method of measuring second-language acquisition *Modern Language Journal* (St Louis, Mo), **58**, 5/6 (1974), 245–50.

Developmental sentence scoring (DSS) is a procedure for estimating a child's language development (mother tongue). Points are given for correct usage and range over eight different categories (eg pronouns, verbs) [scoring described]. The DSS technique was applied to research data on the second-language acquisition of Mexican-American children in traditional and bilingual schools, to see whether DSS would confirm the result of error analysis and whether it would reveal similarities and differences between first- and second-language acquisition. [Results.] It was concluded that many mistakes made in both processes are similar, but documentary similarity between the processes depends on the conceptual framework used to describe the first-language process. DSS is at least partially applicable to second-language learning processes because it measures increased syntactic complexity and greater correctness.

100. **POLITZER, R L, M R Hoover,** and **D Brown**
A test of proficiency in black standard and nonstandard speech *TESOL Quarterly* (Washington, DC), **8**, 1 (1974), 27–35.

As part of the development of a test battery to determine proficiency in

black standard and non-standard speech, a test was devised consisting of a repetition task in which 15 sentences in black standard and 15 in black non-standard English were to be repeated. The sentences were contained within two similar stories recorded on tape by a bidialectal speaker. A black experimenter administered the test individually to 35 black kindergarten children in 1972. The experimenter stopped the tape after each sentence containing a test item and asked the child to repeat the sentence. The response was scored as correct if the child repeated the test item exactly as modelled on the tape. Scores indicated a general balance between standard and non-standard. Subjects were also assigned a balance score, which measured the dominance of non-standard over standard. Scores of the same students on the Stanford Achievement Test and its subsection on letters and sounds correlated positively and significantly with the standard section of the test. Where there was an imbalance in favour of non-standard, there was a significant negative correlation with the Stanford Achievement Test.

101. PORTER, D

Modified cloze procedure: a more valid comprehension test *English Language Teaching Journal* (London), **30**, 2 (1976), 151–5.

[Advantages and weaknesses of cloze procedure (CP) tests are discussed.] The main disadvantage of CP is that the student must not only comprehend but produce the required item. A modified CP provides a set of words for each gap and thus does not demand language production, merely recognition of the correct word. [The choice of suitable distractors is discussed.] [Example; references.]

102. RAASCH, Albert

Das VHS-Zertifikat Französisch: Objektivierte Leistungsmessung im Französischunterricht für Erwachsene [The 'Folk High School' certificate in French: objective measurement of performance in French teaching to adults] *Neusprachliche Mitteilungen* (Berlin), 4/1971, 222-37.

The VHS certificate in French testifies to the achievement of practical basic skills in ordinary language and is a standardised supplementary professional qualification acceptable for leaders taking parties to France. [Descriptions of the tests for 1969 to 1971 are set out with the marking schemes for both oral and written work; the topics for the sub-tests for written work are given. There are examples of the comprehension tests, both written and oral, and the analysis of the results of the 1970 tests is dealt with in some detail. There are comments on possible further developments in examining techniques.]

103. ROBINSON, Peter
Towards a basic procedure in the composition of second language tests *IRAL* (Heidelberg), **8**, 1 (1970), 49–58.

The article is concerned with tests to evaluate aptitude to learn another language, categorise what has been learnt and diagnose error and difficulty, classify students into homogeneous groups, and predict and reveal progress. A brief analysis shows what points each test is intended to cover. Selection of material for the tests will be directly related to the level of the student. An objective or non-objective form for students' answers has to be chosen. The name 'objective' is considered a misnomer as an objective test may not be more reliable than a non-objective type. The most pertinent criticism of multiple-choice objective tests is that the resulting evaluation is entirely passive. The grading of difficulty of the questions and the ordering of them are also determined by the make-up of the group and by the situation in which the language will be used. Correction and standardisation procedures are noted and also the possible application of the test results.

104. ROBINSON, Peter
The composition, adaptation, and choice of second-language tests *English Language Teaching* (London), **25**, 1 (1970), 60–8.

The various kinds of tests available are described and categorised, and the author shows how they can be used and what purpose they can serve. He outlines the basic procedure to follow in selecting, adapting and composing tests.

105. ROBINSON, Peter
Oral expression tests: (1) *English Language Teaching* (London), **25**, 2 (1971), 151–5.

Oral expression tests are neglected, as the subject's answers cannot be treated objectively. The results of a so-called objective test are no more reliable than those of a subjective test. By limiting the subject's freedom of expression, the objective test becomes a tabulation of what the subject knows, whereas the subjective test, which encourages the subject to express himself, evaluates the use he makes of that knowledge.

An oral comprehension test can be either participating or non-participating, but there are many subcategories. [The author characterises tests of reading and oral comprehension from the point of view of the degree to which the student participates.]

The subject's production should be evaluated in four steps: identification, description, classification, and written and numerical assessment of his language competence in a group of situations. The evaluation procedure should reflect the reality of the act of communication.

In the non-participating form the language used is selected by the test

designer. The subject's selection of the appropriate answers is no guarantee that he can use correct forms in speech and writing. The non-participating test is administratively convenient, but not easy to compose; and it is made with a particular group of subjects in mind.

The problems of composing oral participating tests are less complex. Such tests indicate how much of the subject's knowledge of the language is organised for active expression. Non-participating tests should be used for inventory and diagnostic purposes only.

106. **ROBINSON, Peter**
Oral expression tests (2) *English Language Teaching* (London), **25**, 3 (1971), 260–6.

The test situation must be motivating and real, so that the subject speaks freely and so that his production is representative of his general capacity to speak in most situations. The test should be divided into two sections: interview and discussion. Since oral expression involves both question-and-answer and narration, both sections are necessary. The interviewer's subjective impressions should not be neglected. The criteria of comprehensibility and acceptability should underlie any evaluation. [The author illustrates problems of evaluation and gives examples of test items.] The analysis of oral production should be standardised, and student record-cards drawn up.

Too much attention has been given to non-participating (objective) tests. Research should be concerned with analysing and assessing language production, and with the most effective test situation to encourage such production. Non-participating tests evaluate the subject's passive knowledge of the language. The administrative convenience of non-participating tests conceals their limitations. The most effective means of discovering how much a student has benefited from an oral language programme is to place him in a situation where he has to communicate. [Bibliography.]

107. **ROWE, H M**
Language testing at universities *Babel* (Melbourne), **5**, 2 (1969), 14–17.

Foreign-language teaching at university level should aim at complete communication with the native speaker, and testing should reveal the extent to which the examinee has acquired a set of habits that corresponds to the competence of a native speaker. The familiar testing devices of prose composition, essays, and prepared lecturettes should be examined to see exactly what they test. There should be consistency between teaching and testing methods. If the aim is comprehensive use of the foreign language, the examination should be comprehensive which will usually mean an increase in the time given to the testing of audio-lingual skills. Tests should be as objective as possible. Specific aspects to be tested should be assessed as separate items. Though ability in isolated items does

not indicate overall ability to handle the language there must come a stage when the student can handle all aspects such as phonology, morphology, syntax, stress and pitch patterns at once and a certain proportion of marks may go to overall impression as well as to specially selected language points. Too much in university language teaching is still tradition-bound and haphazard.

108. ROY, Robert R
Rate of output – a factor of oral proficiency *Canadian Modern Language Review* (Toronto), **26**, 1 (1969), 14–22.

An investigation into the rate of output by people learning French as a second language is used to clarify the nature of oral proficiency. The method of testing and evaluation is discussed.

109. RUDD, Elizabeth
Language tests for immigrant children *Multiracial School* (London), **1**, 1 (1971), 26–9.

The need for language tests for immigrant pupils is being met by tests now under development at the National Foundation for Educational Research in England and Wales. A diagnostic test seemed the most obvious need but its production would prove an enormous task as a battery of tests would have to be developed to indicate errors and difficulties for immigrant groups from different national backgrounds. Looking at the immediate language-learning goal – English for school purposes – at least delimits an area to be tested.

The concept of language proficiency as comprising listening comprehension, speaking skill, reading and writing is familiar but inadequate to describe English for classroom purposes. One needs to know what kinds of material are being listened to and read. In specifying the nature of the skills to be tested one has to think first of function and only then of actual language usage and the linguistic implications. One needs to know the distance between the child's present command and what his education demands of him. Visits to a large number of first-year junior-school classrooms and a collection of tape recordings of lessons indicated that the sort of questions asked, and the kind of conceptual interests and the grammatical features used to describe them, were identifiable. Vocabulary will be limited as the immigrant child will not have home and play words and differences of background experience between immigrants and native speaker continue to be significant even as the child becomes more bilingual. Finally, care must be taken to use the results of a test only in the way in which it was intended. There is also the problem of cultural bias and the only defence here is a continual alertness to the problem.

110. **SAKO, Sydney**
Writing proficiency and achievement tests *TESOL Quarterly* (Washington, DC), **3**, 3 (1969), 237–49.

A description is given in this article of suitable test items designed to measure listening, speaking, reading and writing skills. The use of these different items will depend upon the objectives of the training course. Teachers of English to speakers of other languages will play a vital role in checking these objectives and needs and finally in constructing the proper instrument for evaluation.

111. **SCHEIBNER-HERZIG, Gudrun,** and **Angela Pieper**
Untersuchungen zur Struktur der Fremdspracheneignung bei neun-jährigen Schülern [Investigations into the structure of the foreign-language aptitude of nine-year-old students] *Linguistische Berichte* (Braunschweig), **39** (1975), 69–79.

In an attempt to predict performance in foreign-language learning, two tests were administered to 55 nine-year-old German elementary school pupils at the beginning and end of their first year of English instruction. The first one was the German version of Carroll's Modern Language Aptitude Test [5 subtests listed]. The test used at the end of the school year was an informal skill test consisting of 7 subtests [listed]. The subtests of the two tests were then correlated using the rank correlation coefficient according to Spearman. The rank of pupils in the group was established by the correlation values: high values tending to unity mean a favourable prediction of achievement, values around zero mean no prediction, and negative values indicate contradiction in the correlated subtests. [Detailed analysis is given of a matrix illustrating correlations as well as of predictions obtained from it.] Using partial rank correlation according to Johnson can make the results more precise. Success in some skills can be predicted more easily than in others. It also appears that in many cases more than one ability is involved in the prediction of performance in one skill.

112. **SEELYE, H Ned**
Field notes on cross-cultural testing *Language Learning* (Michigan), **16**, 1/2 (1966), 77–85.

The author gives examples of questions used experimentally during the development of an objective test of the cross cultural awareness of Americans after residence in Guatemala. He discusses the variables of age, social class, sex, and residence, as well as problems of administering and validating the questionnaire. The questions were pre-tested with both Guatemalan and American groups. The variables of sex and social class were controlled.

113. SEWARD, B H
Measuring oral production in EFL *English Language Teaching Journal* (London), **28**, 1 (1973), 76–80.

Most tests of speaking proficiency tend to be subjective, and are time-consuming to administer or score. There is no general agreement as to what constitutes speaking proficiency or what level of such proficiency should have been reached at a particular stage of general achievement; nor do we know against whom the student's proficiency should be measured. It is possible to measure speaking ability in a foreign language indirectly. The English-speaking ability of students at the American University in Cairo was evaluated by their teachers, and matched with scores obtained on tests of general proficiency in English and listening comprehension. The fairly high correlation obtained gave support to the hypothesis that indirect means can be used to measure a student's oral proficiency. [A critical description is given of the teachers' evaluation procedures.] Teacher evaluations, although they are less effective than objective written tests in measuring speaking proficiency, should not be entirely disregarded, since in individual cases they can be used along with other tests.

114. SHEARER, Eric
A restandardisation of the Burt-Vernon and Schonnell Graded Word Reading Tests *Educational Research* (Windsor, Berks), **18**, 1 (1975), 67–73.

Although word recognition tests continue to be widely employed by teachers, the application of out-of-date norms and varying methods of administration and scoring seriously reduce the usefulness of these tests. Largely as a result of approaches from teachers, it was decided to undertake a large-scale restandardisation of the Burt-Vernon and Schonell tests to produce reliable norms and to standardise administration and scoring procedures. The opportunity was also taken to record the data so as to enable the production of revised orders of words on both tests to reflect contemporary usage, and to relate reading attainment, as assessed by the tests, to three important factors – sex, school organisation and socio-economic status.

115. SKÁLOVÁ, Eva
Návrh na testování anglických mluvnických znalosti na střední škole [A proposal for testing the knowledge of English grammar in secondary schools] *Cizí jazyky ve škole* (Prague), **15**, 9 (1971/2), 401–5.

A set is discussed of specimen tests to be applied to students of English in some grades of the basic nine-year school and of the general (higher selective) secondary school in Czechoslovakia. These tests are not standardised attainment tests, but only proposals which the teachers are expected to modify and adapt to their teaching conditions. The purpose of these tests is to assess half-yearly or annually to what extent the learners

have achieved the objectives in the foreign-language programme. The tests should be elaborated by a group of teachers and a native English speaker should be consulted. Advice is given on how to proceed with the item analysis (for instance, how to determine its facility and discrimination value). The preparation of at least two versions of the test is recommended in order to prevent older pupils who have taken the test previously from influencing younger ones. These forms of grammar testing are suggested: (1) item substitution, (2) item transformation (for example, singular into plural), (3) multiple-choice answers. Three short specimen tests in English grammar for pupils of the eighth grade of the basic nine-year school are given.

116. SMITHIES, Michael
Two methods of testing comprehension *English Language Teaching Journal* (London), **30**, 2 (1976), 144–51.

Two approaches to reading comprehension are described which aim to make students give serious consideration to linguistic details such as number, tense, etc. The first consists of a fairly technical text with a series of multiple-choice questions devised with the learners' particular difficulties in mind (in this case, Thai students) [many examples]. Results showed that there were few errors in the vocabulary section but as much as 92 per cent error with phrases or function words. With practice, students learned to analyse meaning more carefully. In the second method the different component sections can be mixed or separated in a different way [examples]. The methods effectively teach, as well as test, comprehension.

117. SPENCER, Richard E
The influence of disc or tape language laboratory equipment on foreign language speaking test scores *Modern Language Journal* (St Louis, Missouri), **50**, 4 (1966), 207–8.

Eighty-eight students of French were given the MLA Speaking Test, administered on discs, and another 372 were given the same test on tape. Results were scored in common without the scorers knowing which system had been used by individuals. A 10 per cent sample was rescored for reliability checks. There was no significant difference in test scores resulting from the method of recording used.

118. STEIN, Oswald
Was prüfen wir eigentlich? Überlegungen und Vorschläge zur mündlichen Prüfung im Abitur [What are we really testing? Thoughts on, and proposals about, the oral examination for the Abitur] *Praxis* (Dortmund), 4/1972, 357–65.

At present the oral examination of foreign languages for the *Abitur*

requires the pupil to read aloud a passage (which he is allowed a quarter of an hour to prepare) then give a short talk on it or the gist of it in his own words; sometimes this is followed by sentences to translate and questions on grammar; in the time left there is supposedly a conversation, the whole usually before a panel of three examiners. Dissatisfaction with these arrangements led to a small-scale experiment in a Frankfurt grammar school. It was based on the following principles: oral examining should be used to test only what cannot be tested in writing; in an oral examination forming part of a school-leaving examination there should be no testing of skills which have not been practised in the classroom; foreign-language texts in an oral examination should be used solely to stimulate speech activity and should not be regarded as examinable in themselves. The experiment used an assessment table with the following headings: reading (intonation, quality of sound, pronunciation), pupil's short talk (pronunciation, vocabulary, prepositions, grammatical correctness, sentence construction, fluency, development of theme), conversation (as for the short talk, minus pronunciation). The allocation of marks to each subheading was indicated and the total (maximum 100) was then translated into the usual six-point scale for the *Abitur*, with a 5 per cent easier scaling for pupils who were not in the languages stream. Candidates felt more assured that they had been fairly assessed, and the success of the experiment calls for more work on this topic.

119. STOLDT, Peter H

Kontrolle mündlicher Sprachfertigkeiten in Unterricht und Prüfung [The assessment of oral linguistic skills in the classroom and examinations] *Praxis* (Dortmund), **20**, 1 (1973), 33–42.

The methods used in the oral examination in English of the *Volkhochschule* certificate for adults could equally well be applied to school children. Three oral skills are distinguished: (1) *Discrimination* – ways of testing the ability to differentiate between similar-sounding words or sentences. (2) *Aural/listening comprehension* – the pupil distinguishes between pairs of homophones or responds by actions to simple commands. He sits a series of aural comprehension tests with pictures or printed multiple-choice answers. At a higher level more elaborate tapes may be used, but care should be taken to avoid creating tests of memory or of specialised knowledge. Pre-testing the questions individually can help to obviate these dangers. (3) *Oral production:* (a) *Reading aloud* – which should not be judged by general impression but marked for specific points; three types of test are described. It is essential to keep the assessment of reading aloud separate from the assessment of other oral skills. (b) *Oral production* – various types of guided conversation are described, some of which use pictures. These methods are more objective than the use of free conversation; but it is felt that the latter should be part of every oral test, if only for its 'back-wash' effect on teaching methods. [Numerous examples given. Bibliography.]

120. **STUBBS, John Bartow,** and **G Richard Tucker**
The cloze test as a measure of English proficiency *Modern Language Journal* (St Louis, Mo), **58**, 5/6 (1974), 239–41.

A cloze test was administered as an integral part of the English Entrance Examination (EEE) for the American University of Beirut, which consisted of five parts: structure, vocabulary, miscellaneous tests of language aptitude, reading comprehension and the cloze passage. The sample consisted mainly of native speakers of Arabic. The test was scored twice, for exact responses and for contextually acceptable responses (the latter to test whether local English teachers could score them reliably). [Table of results.]

Both scoring methods yielded a high positive correlation with the total EEE scores; this indicates that a cloze test scored for exact replacement could be used by a non-native teacher of English. The cloze-exact scores correlated best with the sub-tests of reading and miscellaneous aptitude; the cloze-acceptable scores correlated best with miscellaneous aptitude. Average difficulty level of each item and average item-discrimination level were both as acceptable as comparable data for the standardised EEE. The most difficult parts of speech were found to be conjunctions and prepositions. From this it is claimed that the cloze technique is a powerful and economical measure of English language proficiency as well as a useful diagnostic tool for the teacher. [References.]

121. **THOMAS, R Murray**
Testing second-language listening skills via television *Audiovisual Instruction* (Washington, DC), **20**, 5 (1975), 24–7.

Instructional television was used in the development of a system for testing English-language listening comprehension among American Samoan pupils in grades 3, 6, 9 and 11. American English is taught as the official second language [background to the language programme, including the provision of an instructional television system in all classrooms]. The need was for group testing of large numbers of pupils in a form which permitted rapid, accurate scoring. Typical problems are the pronunciation of the non-native-speaking teacher and his tendency to give his pupils unwarranted help with the tests. The use of different native speakers has the disadvantage of producing a test situation which is not standardised. Videotape provided the solution – a native speaker in the television studio can present test sentences and the session can be videotaped for broadcasting to the schools. The conducting of the test programme is described: (1) the form of the test; (2) selection of test sentences; (3) planning the layout of test items in the pupils' answer booklets; (4) videotaping the oral presentation and accompanying pictorial test items; (5) administration of the tests, and (6) computerising the answers from the booklets and analysing results. [Pictorial test items illustrated.]

122. TOWNSON, Michael
Testing oral skills at university level *English Language Teaching* (London), **27**, 2 (1973), 199–205.

Attempts made at the University of Erlangen to improve the oral examinations in English are described. Free conversation between examiner and student was the traditional form of test and its 'naturalness' was felt to be more important than the fact that it was hard to score. To overcome the subjectivity of the test, certain safeguards were introduced. An evaluation sheet, listing all the features to be observed and providing a rating scale for each, was drawn up, a criterion of 'near-nativeness' of speech being adopted. [Details of the sheet are given.] In practice, however, it was found difficult to distinguish fluency from grammar and vocabulary, and between the various levels of auditory comprehension. It was also realised that fluency and speed of delivery are influenced by factors unconnected with command of English. It was decided that the assessment of fluency was sufficiently covered under other headings. A trial examination was conducted to eliminate discrepancies between examiners' judgement, but almost total agreement was found on every test. The problem of converting points into numerical marks has not been solved. There is the question of how much importance should be attached to each feature and also of where to place the pass mark.

123. TRAILL, A
Concerning the diagnosis and remedying of lack of competence in a second language *Language Learning* (Michigan), **18**, 3/4 (1968), 253–8.

A native speaker's linguistic competence cannot be equated with a sample of his performance, since non-linguistic information is contained in the latter and since any sample of performance is fortuitous and finite. The bilingual's lack of competence in the second language may be partial or, in parts of the language, complete. Lack of competence may not be reflected in peformance. [The author describes a test to discover what a young African child knows about 'nested structures' in English.] Command of various other types of structure can be tested in this way. The test diagnoses an unrecognised or unrecognisable lack of competence in the second language and suggests the type of remedial material needed. Reading materials might have to be replaced by graded passages and comprehension tests.

The appreciation of literature depends, in part, on grasp of the linguistic system in which it is expressed.

124. TRUCHOT, Claude
Les tests de langue: réévaluation critique [Language tests: a critical re-evaluation] *Langues Modernes* (Paris), **65**, 2 (1971), 103–12.

The problems which are raised in composing tests are the same as those posed by language teaching methodology in general and the application of

linguistics and psychology to teaching. Care has to be taken that the test measures exactly what it is intended to measure. The essential quality of a test is its validity. Diagnostic tests include progress and achievement tests; proficiency tests are prognostic. Aptitude tests can have a certain value but one may wonder to what extent they judge what has already been acquired rather than what may be acquired. Structural analysis has proved useful in test composition because it has enabled language to be analysed but such a system is limited because it represents the elements of an abstract system. Testing a knowledge of structures presupposes that structures can be isolated and taught separately, that they can be memorised by the repetition of correct responses. The linguist, however, should not be expected to compose a pedagogical method from his discoveries. He can only supply hypotheses which the teacher can verify in practice.

It is really only tests of comprehension which can be evaluated objectively. Where creative powers are being tested one can only attempt to establish rigorous criteria for a subjective evaluation. The problem would be simplified if we knew to what extent the measurement of language comprehension could be used to predict mastery of expression. Some recent experiments show a high correlation between a comprehension test and a 'scored interview'. Psychologists have drawn our attention, as far as foreign-language teaching is concerned, to the role of attention and short-term and long-term memory. Linguistic analysis of utterances has revealed a number of variables: syntactic, morphological, phonetic and prosodic complexities; lexical semantic and even rhetorical difficulties. Advanced oral tests can introduce more complex variables such as pauses, hesitations and dialectal variations. Progress tests are simplified in that they are constructed from a corpus. A proficiency test is more complicated. [Details of suggested procedure.]

Evaluating written language is a different technique but can still consist in multiple choice questions on a text. Such a test will have to be based on the lexical, syntactic and stylistic variables which enter into comprehension of a text. Traditional techniques of translation and essay writing are skills in themselves and as a means of evaluating knowledge of a language can only be subjective. One might go so far as to say that there cannot be a method unless there is a reliable, scientific and objective evaluation of its results.

125. **UPSHUR, J A**
Cross-cultural testing: what to test *Language Learning* (Michigan), **16**, 3/4 (1966), 183–96.

Measures of language ability have limited power to predict who will be able to function effectively in a new linguistic and cultural environment. There is a need for test instruments and procedures which can measure cultural understanding reliably. [The author takes Seelye's article *Field notes on cross-cultural testing* as a starting-point (entry 112).] In

cross-cultural testing one is concerned with whether the 'cultural stranger' can behave (non-lingustically) in such a manner that his intended meanings are understood by members of the foreign culture community, and whether he understands their intended meanings when he observes their behaviour or its products. The preparation of such tests involves considerable work. Four aims of testing are recognised: the user wishes to (a) determine how an individual would perform in situations of which the test gives a sample, (b) predict an individual's future performance, (c) estimate an individual's present status on some variable external to the test, (d) infer the degree to which the individual possesses some quality. [The author suggests how these aims can be translated into prescriptions for test construction.]

In some instances the 'cultural stranger' will understand and behave correctly in the foreign culture community by virtue of similarities between his own and the target culture.

A description of the universe of test items must include: (1) the observed behaviour (or behavioural products) to be understood, and (2) the patterns to be appropriately performed. [The author offers a detailed criticism of Seelye's views on Guatemalan and American cultural patterns, and states some further problems of cultural testing which remain to be solved.]

126. **UPSHUR, John A**
Objective evaluation of oral proficiency in the ESOL classroom *TESOL Quarterly* (Washington, DC), 5, 1 (1971), 47–59.

It is appropriate to test when we do not know to what extent teaching has been successful. Classroom testing by language teachers is essential. Testing and its function in teaching is defined by the procedure of giving the student a task, observing his performance, and informing him of his degree of success. Classroom testing, because it is a necessary part of teaching, may give a better indication of student ability than a standardised test.

There is a place for tests not used directly in the course of instruction. A good test should be objective. Objectivity is achieved by means of a multiple-choice format, by 'directed speaking', or by measuring the time taken to perform an oral task. [Various other suggestions are mentioned.]

The first step in the testing process – deciding what to test and how – is usually subjective. The teacher who has kept accurate records of the students' daily oral performance has given an excellent oral test. [A specimen record sheet is given. The author also describes games which can be used as oral proficiency tests.]

If the teacher knows the day-to-day objectives of the course, he can conduct a running test of skills involved in the ability to communicate. If he knows why his students are learning English, he can test their ability to communicate in the kind of situation for which they are being trained.

127. VALETTE, Rebecca M

Improving multiple-choice grammar tests in German *German Quarterly* (Wisconsin), **40**, 1 (1967), 87–91.

A false picture of a student's knowledge may be gained from multiple-choice tests because of the possibility of guessing the right answers. This study was carried out to establish on what grounds the poorer student tended to select the response to an item he did not know. Selected examples are given from a comprehensive objective-type final examination administered to five classes of first-year, first-semester college German students.

In examining the responses it was noticed that the poorer students tended to fill blanks with items possessing an inflectional ending identical to the final consonant and/or vowel of the word preceding the blank. Such items could be eliminated from tests. The inflectional system of German lends itself to such modification, for words of another number or gender can always be found to replace the unwanted option. Suggestions are made for combinations of endings with propositions to be avoided. By eliminating dubious multiple-choice and fill-in-the-blank items, the teacher will take a definite step in improving the reliability of his tests.

128. VOLODIN, N V

Kontrol' ponimaniye inoyazychnogo teksta [Assessing comprehension of a foreign textual material] *Inostrannye yazyki v shkole* (Moscow), **5**,(1972), 29–37.

A foreign-language course should consist not only of teaching materials, presented in accordance with a gradual progression of lexis and syntax from simple to complex, but also of testing materials graded and presented in the same way. The principles on which tests are constructed should take careful account of the learner's ability to recognise forms of language, his skill in using them and the degree to which he has converted such forms into habits. Objectivity must be observed in assessment and tests should be suited to the level of attainment demanded by the teachers at any given point in the course. To satisfy these conditions multiple-choice tests are recommended, divided into three levels of attainment and graded according to the difficulty of the lexis and syntax. At each level the learner should score maximum marks before proceeding. Thus, the construction of such tests must ensure that the learner cannot fail at a lower level but pass at a higher one. [Techniques for improving tests are outlined.] Tables for evaluating performance in each of the three levels may conveniently include a five-point rating, to help both learner and teacher judge progress. [An example, based on two different types of material, partially illustrates one such set of tripartite tests.]

129. WESDORP, H

Het toetsen van de schrijfvaardigheid in het Engels op het schoolonderzoek M.A.V.O. – H.A.V.O. – V.W.O. [The testing of English writing skills in Dutch secondary schools] *Levende Talen* (The Hague), **304**, (1974), 1–19.

The main problem is how to define writing skills more precisely, and how

to test them reliably and validly in schools. The definition of writing skills is taken from De Vries: 'the ability to express oneself in writing, in simple but correct language, on general, non-specialised subjects'. Although examination results can be crucial for the individual, teachers frequently misunderstand the concepts of reliability and validity. Three types of test can be distinguished: (1) the direct method, including composition scales, the analytic schema and the global method using one or more assessors; (2) semi-objective methods, including the interlinear exercise and the Peellandt College blank-filling system; (3) objective methods, including the integrated approach (eg the STEP writing test of Educational Testing Service and the Tests of Academic Progress by Scannel *et al.*) and the sub-skills approach (eg the American Entrance Examination Board test, the SRA Writing Skills Test, the Freyman Content Evaluation Series, the Missouri College English Test, the Test of English Usage of the California Test Bureau, the English Expression Test of the Educational Testing Service, and the College English Placement test). [These different types of test are commented on, with illustrations.] [Bibliography.]

130. ZIERER, Ernesto
The test of aural perception in foreign language teaching from the standpoint of information theory *IRAL* (Heidelberg), **9**, 2 (1971), 125–30.

Although it is generally accepted that the correct production of the sound segments of a language presupposes their correct recognition, the reverse is not true. Suitable recognition tests have been elaborated by various scholars. A standard aural test of perception in a certain language for students of a different language background is usually obtained after a repeated try-out with a homogeneous group of students. The results obtained, expressed in values of information theory, may help to determine the quality of the test.

131. ZIRKEL, Perry A
A method for determining and depicting language dominance *TESOL Quarterly* (Washington, DC), **8**, 1 (1974), 7–16.

A practical model for determining and depicting language dominance, given the general nature and needs of bilingual education programmes in the United States, is presented. The use of parallel tests of aural ability can give an initial indication of the language dominance of children who might otherwise be classified by surname as 'Spanish-speaking' or 'bilingual'. The results can be organised for placement or programmatic purposes into a continuum, from Spanish- to English-dominance. However, this one-dimensional representation does not permit differentiation of ability level within each language. For example, a 'balanced alingual' child, who is equally deficient in each language, is represented at the same point on the

continuum as a 'balanced bilingual' child, who is natively proficient in each language. A two-dimensional representation is therefore proposed, elevating the left side of the continuum into a vertical plane so as to form a two-dimensional matrix. This model aids the effective planning of programmes according to bilingual dominance and proficiency. Several such linguistic and programmatic patterns are discussed. [References.]

132. **ZYDATISS, Wolfgang**
Some test formats for elicitation procedures *IRAL* (Heidelberg), **12**, 4 (1974), 281–7.

It is suggested that not only the 'intermediate language' of individuals but also of particular groups of learners should be studied, since the findings could be more readily exploited in second-language teaching. Elicitation procedures are necessary to make learners generate certain structures or shades of meaning which conventional error analysis fails to elicit. Three types of test are compared: (1) translation of non-contextualised German sentences (mother tongue) into English, (2) translation of non-contextualised German sentences into English under specific surface constraints, and (3) a multiple choice test which could have more than one correct answer (thus indicating what the learner considers 'acceptable'). Heuristic methods such as error analysis and contrastive analysis should be complementary; the development of elicitation procedures becomes a cyclic process.

4. Books

Bibliographies

B1. **BUROS, Oscar K,** *editor*
Seventh mental measurements yearbook, Vol 1 Gryphon Press, New Jersey, 1972. 935pp.

Pages 536 to 606 list published tests in foreign languages and provide authoritative reviews of their contents and construction.

B2. **FEHSE, K D** and **Wulf Praeger**
Bibliographie zum Testen in der Schule: Schwerpunkt: Fremdsprachenunterricht Becksmann, Freiburg i. Br., 1973. viii, 235pp.

Lists numerous books and published articles (with many annotations) on theory and construction of tests, and 52 published tests (various languages). Text in German.

B3. **NOSTRAND, Howard L,** and others
Research on language teaching: an annotated bibliography 1945–64 2nd edn. rev. University of Washington Press, Seattle and London, 1965. xxiv, 373pp.

Pages 173 to 184 and 216 to 226 cover research on language testing. Valuable record of past development.

B4. **SAVARD, J G**
Analytical bibliography of language tests Presses de l'Université Laval, Quebec, 1969. xviii, 372pp.

Brief descriptions of a variety of second-language, mother-tongue and bilingual tests in English, French, German, Italian, Spanish (some unpublished and many now out of date).

B5. **WALTERS, Theodore W**
The Georgetown bibliography of studies contibuting to the psycholinguistics of language learning Georgetown University Press, Washington DC, 1965. ix, 125pp.

Provides references to psychological research related to language testing up to 1964.

Test design and construction

B6. **BURSTALL, Clare,** and others
Primary French in the balance National Foundation for Educational Research in England and Wales, Windsor, 1974. 304pp.

See especially pages 247 to 292 'Notes on the development of the measures of attitude and achievement, which reviews many of the problems of language testing and provides a summary of recent research.

B7. BURSTALL, Clare
'The main stages in the development of language tests' in Stern, H H, *(editor) Languages and the young school child* Oxford University Press, 1969. viii, 270pp. (Language and Language Learning 24).

Short description of construction of tests for younger children.

B8. CARROLL, John B
'The prediction of success in intensive foreign language training' in Glaser, R, *(editor) Training research and education* University of Pittsburgh Press, 1962. pp 87 - 136.

Establishes some general principles of aptitude testing.

B9. CLARK, John L D
Foreign language testing: theory and practice Center for Curriculum Development, Philadelphia, 1972. x, 174pp.

Practical handbook on prognostic, achievement and proficiency testing. Examples from French, Spanish, etc.

B10. DAVIES, Alan
'Aptitude for and proficiency in French in the first year of the UK secondary school' in Perren, G E, and J L M Trim, *(editors) Applications of linguistics* Cambridge University Press, 1971. pp. 177 - 179.

Brief review of problems of establishing an effective measure of language aptitude in schools.

B11. DAVIES, Alan, *editor*
Language testing symposium: a psycholinguistic approach Oxford University Press, 1968. viii, 214pp.

Eleven papers on aspects of language testing and associated psychometric and linguistic problems. Examples mostly from English as a foreign language.

B12. DYSON, A P
Oral examining in French Modern Language Association, 1972. 50pp.

Reports research carried out on problems of design and construction of reliable oral tests at CSE and GCE 'O' levels.

B13. GREEN, P S, *editor*
The language laborary in school: performance and prediction: an account of the York study Oliver & Boyd, Edinburgh, New York, 1975. (viii), 242pp.

Includes a section on testing (P S Green) which describes the assembly of an aptitude test battery and of attainment and proficiency tests in German for use at secondary level, and another section on the interpretation of test results.

B14. HARRIS, David P
Testing English as a second language McGraw-Hill, New York, 1969. viii, 151pp.

Handbook on design, construction, administration and interpretation of tests.

B15. INGRAM, Elisabeth
'Language testing' in Allen, J P B, and S P Corder, *(editors) The Edinburgh course in applied linguistics, Vol 3* Oxford University Press, 1974. (Language and Language Learning 35) pp. 313 - 343.

Brief description of principles of test construction with examples.

B16. JONES, RL, and B Spolsky, *editors*
Testing language proficiency Center for Applied Linguistics, Washington, DC, 1975. v, 146pp.

Papers on testing speaking proficiency, listening comprehension and reading comprehension which examine theoretical and practical problems.

B17. LADO, Robert
Language testing: the construction and use of foreign language tests Longmans, 1961. xxiii, 389pp.

Provides many examples (mostly from Spanish and English) of techniques of test construction. Clear introduction for the non-specialist, although now dated in some aspects.

B18. MACKEY, William F
Language teaching analysis Longmans, 1965. xiv, 554pp.

Pages 403 to 416 describe, with examples, the purposes of various types of test.

B19. MODERN LANGUAGE ASSOCIATION OF AMERICA
A handbook on foreign language classroom testing: French, German, Italian, Russian, Spanish by André Paquette and Suzanne Tollinger. Modern Language Association of America, New York, 1968. x, 227pp.

Designed for the needs of the classroom teacher who requires to make tests to support a teaching programme.

B20. PARK, B, and J D Simpson
French Methuen Educational, 1974. xxii, 74pp. (Handbooks on Objective Testing).

Introduction to principles of objective testing followed by detailed information on construction of tests of listening and reading comprehension in French. Many examples related to CSE and GCE 'O' levels.

B21. RIVERS, Wilga M
Teaching foreign-language skills University of Chicago Press, London and Chicago, 1968. xii, 404pp.

Pages 286 to 317 review principles on which language tests may be based.

B22. VALETTE, Rebecca M
Modern language testing: a handbook Harcourt Brace & World, New York, 1967. xvii, 200pp.

Designed to assist teachers to make their own tests: practical examples from French, German, Italian and Spanish.

Examples of published tests

B23. CARROLL, John B, and S M Sapon
Modern language aptitude test Psychological Corporation, New York, 1959.

Designed primarily to assess the aptitude of adults for foreign language learning *(Modern language aptitude test – elementary* Psychological Corporation, New York, 1967, is a 'junior' version for schoolchildren.)

B24. MODERN LANGUAGE ASSOCIATION OF AMERICA
Cooperative foreign language tests (French, German, Italian, Russian, Spanish) Educational Testing Service, Princeton, New Jersey, 1964.

Tests at secondary school level covering generalised skills.

B25. MODERN LANGUAGE ASSOCIATION OF AMERICA
Foreign language proficiency tests for teachers and advanced students (French, German, Italian, Russian, Spanish) Educational Testing Service, Princeton, New Jersey, 1962.

For use at advanced level, covering teaching as well as linguistic skills.

B26. PERCIVAL, T S
Basic proficiency in French Hodder & Stoughton, 1976.

A standardised battery of tests for secondary school pupils in listening, reading and writing. British norms.

B27. PIMSLEUR, P
Language aptitude battery Harcourt, Brace & World, New York, 1966.

For use with younger as well as older students.

B28. PIMSLEUR, P
Modern foreign language proficiency tests (French, German and Spanish) Harcourt, Brace & World, New York, 1967.

For use at secondary school level. US norms.

Index of authors and editors

References are to entry numbers

Anderson, Jonathon, 1,2,3
Atai, P, 57
Bachman, Lyle F, 4
Beardsmore, H Baetens, 5,6
Beaudot, Alain, 7
Beneš, Eduard, 37
Blatchford, C H, 8
Bonheim, Helmut, 9
Breitung, H A, 10
Brière, Eugène J, 11,12
Brings, Friedhelm, 13
Brosch, Gerd, 14
Brown, D, 100
Bubeníková, Libuše, 15
Buros, Oskar K, B1
Burstall, Clare, B6, B7
Butzkamm, Wolfgang, 16,17
Carroll, John B, 18,B8,B23
Castigliano, Luigi, 19
Catford, J C, 20
Chastain, Kenneth, 21
Clark, John L D, 22, B9
Clarke, Mark A, 23
Cohen, Arthur M, 24
Connell, Peter A, 25
Conrad, Christine A, 89
Cooper, Malcolm D, 26
Cory, Mark E,27
Cosgrave, Desmond P, 28
Darnell, Donald K, 29
Davies, Alan, 30, B10, B11
Denham, Patricia A, 31
Deyes, A E, 32
Di Cristo, Albert, 33
Dimitrijević, Naum R, 34
Djordjević, Dusan, 34
Doble, Gordon, 35
Donnerstag, Jürgen, 36
Dušková, Libuše, 37
Düwell, Henning, 38
Dyson, A P, B12
Ewen, Elizabeth, 39, 47
Famiglietti, Michele, 40
Farrington, Brian, 41,42,43
Fehse, K D, B2
Feltkamp, H W, 44
Folland, David, 45
Forrest, Ronald, 46
Gipps, Caroline, 39, 47
Gommes-Judge, Anne, 48
Green, P S, B13
Handforth, D A, 49
Hardin, James N, 50
Harris, David P, B14
Harrison, Andrew, 51
Henning, Grant H, 52
Hill, I L, 53
Hill, L A, 54
Hoover, M R, 100
Hutchings, Geoffrey, 55
Ingram, Elisabeth, 56, B15
Irvine, P, 57
Jakobovits, Leon A , 58
James, Carl, 59
Jeanneret, René, 60
Johansson, Stig, 61
Jones, R L, B16
Kamratowski, Joachim, 62
Kankashian, Ibrahim, 63
Kobersky, Eva, 64
Kollárik, Teodor, 65
Kollmannová, Ludmila, 66
Komorowska, Hanna, 67
Künne, Wulf, 68
Lado, Robert, B17
Lasch, Miguel, 69
Leutenegger, Ralph R, 70
Levine, Josie, 71
Löschmann, Marianne, 72
Löschmann, Martin, 72
Lowe, J Edgar, 73
McLeod, John, 74
Mackey, William F, B18
Madson, Harold S, 75
Maman, André, 76
Matthews-Březský, R J H, 77
Modern Language Association of America, B19, B24, B25
Mohan, Bernard A, 78
Möhle, Dorothea, 79

Moller, Alan, 80
Most, R L, 81
Mueller, Theodore H, 70
Nehls, Dietrich, 82
Niżegorodcew, Anna, 83
Nostrand, Howard L, B3
Novák, Zdeněk, 84
Nübold, Peter, 85
Oller, John W, 57, 86, 87, 88, 89, 90
Page B W, 91
Park, B, B20
Pauels, Wolfgang, 92
Pendlebury, A C, 93
Percival, T S, B26
Pfaff, Harald, 94
Pickett, G D, 95
Pieper, Angela, 111
Pimsleur, P, B27, B28
Pinel, Nicole, 96
Plaister, Theordore H, 97
Politzer, Robert L, 98, 99, 100
Porter, D, 101
Praeger, Wulf, B2
Raasch, Albert, 102
Renkin, A, 6
Rivers, Wilga M, B21
Robertson, David, 45
Robinson, Peter, 103,104,105,106
Rowe, H M, 107
Roy, Robert R, 108
Rudd, Elizabeth, 109
Sako, Sydney, 110
Sapon, S M, B23
Savard, J G, B4
Scheibner-Herzig, Gudrun, 111
Seelye, H Ned, 112
Seward, B H, 113
Shearer, Eric, 114
Simpson, J D, B20
Skálová, Eva, 115
Smithies, Michael, 116
Snider, James G, 25
Spencer, Richard E, 117
Spolsky, B, B16
Stein, Oswald, 118
Stoldt, Peter H, 119
Streiff, Virginia, 90
Stubbs, John Bartow, 120
Thomas, R Murray, 121
Townson, Michael, 122
Traill, A, 123
Truchot, Claude, 124
Tucker, G Richard, 120
Upshur, J A, 125, 126
Vallette, Rebecca M, 127, B22
Volodin, N V, 128
Walters, Theodore W, B5
Wershow, Irving R, 70
Wesdorp, H, 129
Zierer, Ernesto, 130
Zirkel, Perry A, 131
Zydatiss, Wolfgang, 132